APPENDIX E. COMPARABILITY OF FRM FINE MASS MEASUREMENTS WITH RECONSTRUCTIONS OF FINE MASS FROM PM$_{10}$ COMPONENTS

This appendix summarizes the approach used for reconstructing fine mass concentrations from measurements of sulfate, nitrate, and carbon. Three reconstructions were developed. Each uses measurements of sulfate and nitrate from PM$_{10}$ samples. An estimate of total (organic plus black) carbon was added to the sum of sulfate and nitrate, with the different reconstructions differing in the way that total carbon was estimated.

For a limited number of PM$_{10}$ samples, total carbon was measured. These measurements were compared with both coefficient of haze (CoH) and carbon monoxide (CO) measurements. CoH is a measure of light absorption, which largely depends upon levels of black carbon. Because black carbon and organic carbon concentrations are typically well correlated, measurements of CoH potentially provide a means for estimating total carbon levels. The CoH database is extensive (Table E1). Not all sites shown in Table E1 were used for reconstructing fine PM mass concentrations; however, many of the sites listed had measurements of PM$_{10}$ sulfate and nitrate, which we then combined with the estimated carbon concentrations.

Table E1. Number of sites in each air basin with coefficient of haze data, by year.

YEAR	AIR BASIN																
	GBV	LC	LT	MD	MC	NCC	NC	NEP	SV	SS	SD	SFB	SJV	SCC	SC	MEX	OUT
1980	0	2	0	2	1	1	1	0	10	1	0	22	8	4	8	0	0
1981	0	2	0	0	2	1	0	0	10	1	0	16	11	4	6	0	0
1982	0	1	0	0	1	1	0	0	15	0	0	14	10	4	2	0	0
1983	0	1	0	0	1	2	0	0	13	0	0	13	14	8	0	0	0
1984	1	1	0	0	1	2	0	0	14	0	0	13	12	7	0	0	0
1985	1	1	0	0	1	2	1	0	13	0	0	14	6	7	0	0	0
1986	1	1	0	0	3	2	1	0	12	0	0	14	6	6	0	0	0
1987	1	1	0	0	3	2	0	0	13	0	0	13	8	7	0	0	0
1988	1	1	1	0	1	2	1	1	12	0	0	13	8	7	0	0	0
1989	1	1	1	0	2	2	0	1	14	0	0	14	7	6	0	0	0
1990	1	1	1	0	2	2	0	1	13	0	0	14	9	6	0	0	0
1991	1	1	1	0	1	1	0	0	14	0	6	14	8	6	0	0	0
1992	0	1	2	0	1	1	0	0	12	0	4	14	6	5	0	0	0
1993	0	1	1	0	0	1	0	0	13	0	4	13	6	5	0	0	0
1994	0	1	1	0	0	1	0	0	12	1	4	13	7	5	0	0	0
1995	0	1	1	0	0	1	0	0	11	1	4	13	6	5	0	0	0
1996	0	1	1	0	0	0	0	0	12	1	4	7	6	5	0	0	0
1997	0	0	1	0	0	0	0	0	11	1	4	7	6	5	0	0	0
1998	0	1	1	0	0	0	0	0	11	1	0	7	6	5	0	0	0
1999	0	1	1	0	0	0	0	0	11	1	0	7	6	5	0	0	0
2000	0	1	1	0	0	0	0	0	11	1	0	7	6	6	0	0	0
2001	0	1	1	0	0	0	0	0	11	0	0	7	6	4	0	0	0
2002	0	1	0	0	0	0	0	0	11	1	0	6	6	3	0	0	0

We estimated total carbon from CoH as:

E1.   C from CoH = $3.4*(1/10)*71*CoH^{0.76}$

The factor 3.4 is an approximate mean ratio of total to black carbon observed in California PM samples. The factor 1/10 is the inverse of the black carbon absorption efficiency, 10 m^2/g (which is a commonly used factor for converting from units of mass concentration in μg m^{-3} to light extinction in inverse megameters). The remaining terms convert from the reported CoH units (soiling index) to inverse megameters, according to CARB guidance. Comparisons of measured PM_{10} total carbon with total carbon calculated according to Equation E1 showed good agreement (Figure E1). We set our calculated carbon concentrations to be missing if the calculated values were physically impossible according to either of the following conditions:

E2. $C > PM_{2.5}$ mass $- SO_4 - NO_3 - NH_4$
       $C > PM_{10}$ mass $- SO_4 - NO_3 - NH_4$

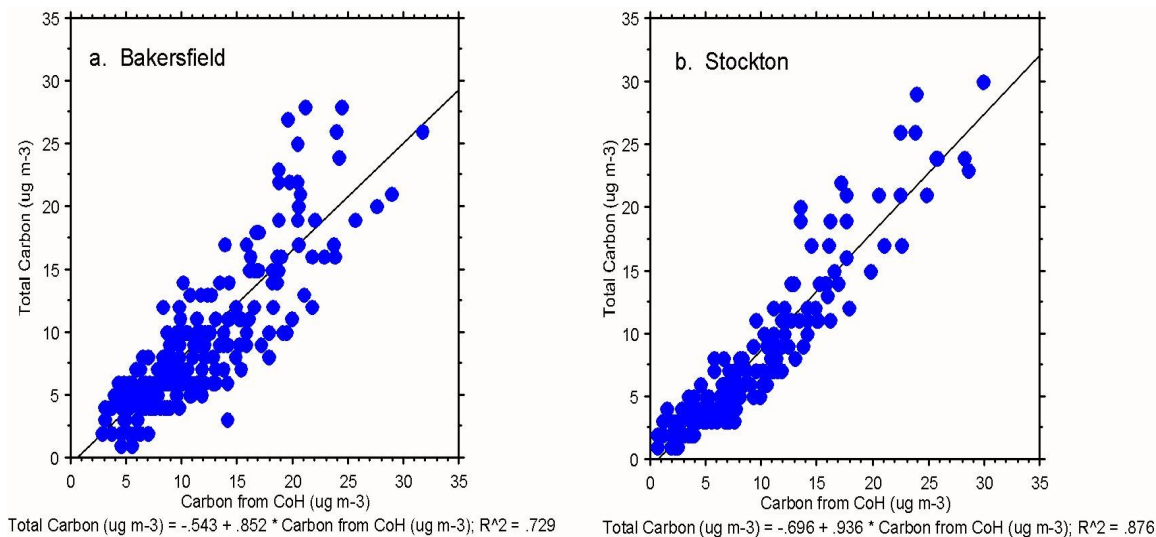If NH4 was not measured, we used $NH4 = (1/2)*(18)*[(SO_4 /96) + (NO_3 /62)]$.

Figure E1. Comparisons of measured total carbon with carbon estimated from CoH.

Equation E2 is a conservative criterion, since PM mass includes geological material as well as mass associated with organic carbon compounds but not included within the measurement of carbon itself (i.e., atoms of oxygen and hydrogen). We also compared measured total carbon with CO concentrations (Figure E2). Combining data from all sites, the generic estimator was

E3.   total carbon (in $\mu g\ m^{-3}$) = 0.008 * CO (in ppbv).

In applying Equation E3, we used the criteria specified in Equation E2 to exclude physically impossible estimates.

PM10 Carbon (ug m-3) = 2.849 + .008 * CO (ppbv); R^2 = .744

PM10 Carbon (ug m-3) = -1.582 + .01 * CO (ppbv); R^2 = .762

PM10 Carbon (ug m-3) = -.049 + .009 * CO (ppbv); R^2 = .669

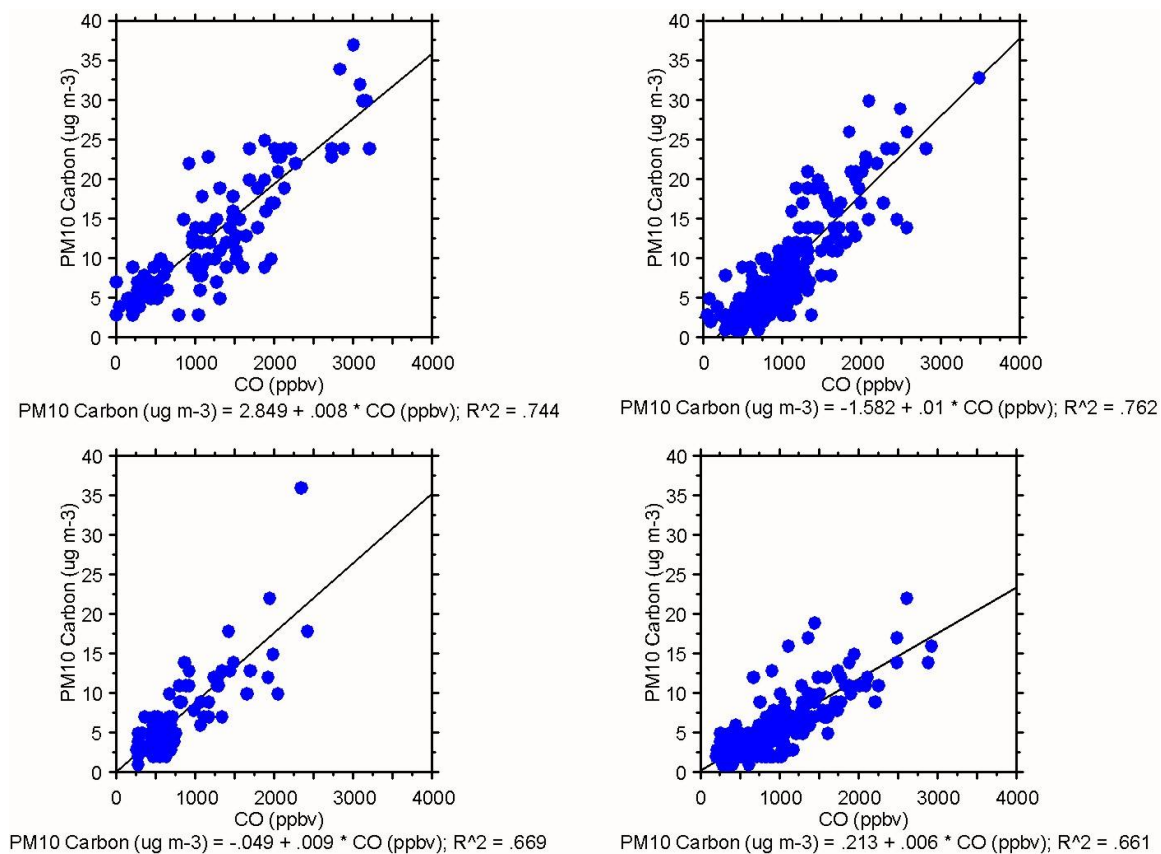PM10 Carbon (ug m-3) = .213 + .006 * CO (ppbv); R^2 = .661

Figure E2. Comparisons of measured total carbon with carbon monoxide.

Estimates of fine mass concentrations were constructed from the sum of sulfate, nitrate, and total carbon, with total carbon values derived from either measurements, CoH, or CO as described above. Additional mass would normally be associated with each of these components since sulfate and nitrate typically occur as partially or fully neutralized ammonium compounds. Organic mass includes organic carbon plus associated elements, such as hydrogen and oxygen. For our purposes, the lack of detailed sample information precluded adjusting the sulfate, nitrate, and total carbon values to more accurately reflect their associated mass contributions, but the lack of adjustment should not bias the relative contributions of these three major components very much. Based upon molecular weight, the ratio of ammonium sulfate to sulfate is 1.375:1, and the ratio of ammonium bisulfate to sulfate is 1.20:1. The ratio of ammonium nitrate to nitrate is 1.29:1. Organic mass is often estimated as 1.4 times organic carbon, though substantial variability exists (our total carbon estimates would include both organic and black carbon, and the latter typically is not associated with additional elements). Thus, each of the three major components should contribute an additional 20 to 40 percent of associated mass to the total fine mass.

Taking the sum of sulfate, nitrate, and total carbon and regressing measured FRM mass concentrations against yielded good ($r^2 > 0.8$) predictors of fine mass concentrations (Figure E3). Using only sulfate and nitrate to predict fine mass concentrations was less reliable and did not meet the criterion of $r^2 > 0.8$ (Figure E3). Figure E3 shows the generic estimators (i.e., regression coefficients developed by lumping all sites together). As discussed in the previous appendices, it was possible to develop reconstructions on a site-specific basis for locations having both FRM and other measurements. We used site-specific estimates if they were based on n>30 comparison measurements and $r^2 > 0.8$. The differences between the site-specific and generic estimates were not large. As previously noted, for both the site-specific and generic predictors, we excluded any predictions for which the predicted $PM_{2.5}$ mass concentration exceeded the measured PM_{10} mass concentration by 10 $\mu g\ m^{-3}$ or more, since such values would be physically impossible and the difference would generally be greater than our estimated uncertainties.
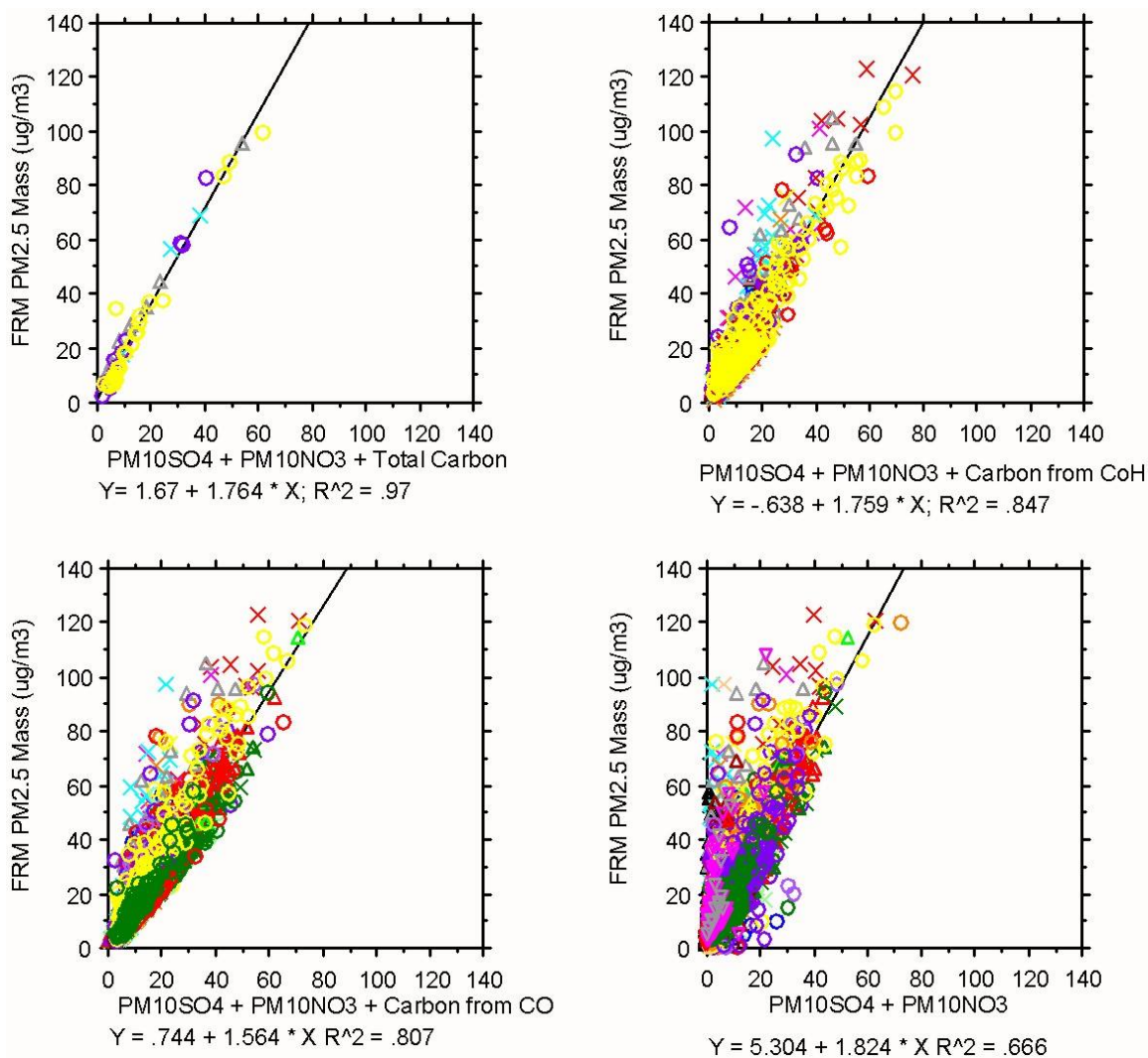
Figure E3. Measured FRM fine mass versus concentrations of the sum of sulfate, nitrate, and total carbon. The carbon concentrations were either measurements made on the PM_{10} samples, or estimated from CoH or CO concentrations. The comparisons of FRM fine mass to the sum of PM_{10} sulfate and nitrate, without including carbon, exhibited lower correlations, $r^2<0.8$.

APPENDIX F. COMPARABILITY OF LIGHT EXTINCTION AND FINE MASS MEASUREMENTS FROM FRM AND DICHOTOMOUS SAMPLERS

Light scattering measurements provide another potential predictor of fine PM concentrations. An extensive set of CARB measurements of light-scattering is available, especially for locations in the Sacramento Valley (Table F1). We investigated the comparability of light scattering measurements made by the CARB with fine mass measurements. Light scattering is measured by nephelometers, which are not size selective, and coarse particles contribute to light scattering, albeit much less than do fine particles. An additional confounding factor is the contribution of fog or cloud droplets, which, when present, tend to cause very high light scattering. We expected that the CARB nephelometers, which are heated, would minimize the fog contribution. We calculated light scattering due to particles (b_{sp}) as:

$$F1. \qquad b_{sp} = (100*24\text{-hour light scattering}) - 10$$

Equation F1 converts the light-scattering data in the CARB database from 10^{-4} meters to 10^{-6} meters, a more convenient unit also known as inverse megameters (Mm^{-1}). Subtraction of 10 approximately removes Rayleigh scattering (light scattering by molecules). Many very large values ($b_{sp} > 1000$) occurred in the data, indicating that fog or cloud droplets likely were affecting the nephelometer values in spite of heating. We excluded samples having $b_{sp} > 800$ Mm^{-1}. For comparison, open (unheated) nephelometer measurements made in the San Joaquin Valley during the IMS95 were less than 500 Mm^{-1} whenever the RH was less than 90 percent and above 500 Mm^{-1} on nearly all occasions with RH exceeding 90 percent (McDade, 1997). Extinction efficiency (light extinction per unit concentration of a chemical component) varies with RH and chemical composition, but is generally in the range of 3 to 20 m^2/g for sulfate, nitrate, and organic carbon (McDade, 1997), so that 800 Mm^{-1} corresponds to approximately 100 to 150 μg m^{-3} fine mass concentration for RH < 90 percent. For comparison, some maximum recorded 24-hour FRM fine mass concentrations in the data base were 87.8 μg m^{-3} at Los Angeles – North Main, 101 μg m^{-3} at Stockton, and 154 μg m^{-3} at Bakersfield.

Table F1. Number of sites in each air basin that have 24-hour light scattering data, by year.

YEAR	AIR BASIN																
	GBV	LC	LT	MD	MC	NCC	NC	NEP	SV	SS	SD	SFB	SJV	SCC	SC	MEX	OUT
1980	0	1	0	0	0	0	0	0	6	0	0	0	0	0	2	0	0
1981	0	1	0	1	0	0	0	0	6	0	0	1	3	0	2	0	0
1982	0	1	0	0	0	0	0	0	10	0	0	1	3	0	2	0	0
1983	0	1	0	0	0	0	0	0	11	0	0	1	2	0	4	0	0
1984	0	1	0	0	0	0	0	0	12	0	0	1	2	0	3	0	0
1985	0	1	0	0	0	0	0	0	11	0	0	1	2	0	5	0	0
1986	0	1	0	0	0	0	0	0	10	0	0	1	3	0	4	0	0
1987	0	1	0	0	0	0	0	0	11	0	0	0	3	0	4	0	0
1988	0	1	0	0	0	0	0	0	10	0	0	0	3	0	4	0	0
1989	0	1	0	0	0	0	0	0	11	0	0	0	3	0	4	0	0
1990	0	1	1	0	0	0	0	0	10	0	0	0	3	0	4	0	0
1991	0	1	1	0	0	0	0	0	9	0	0	0	3	0	4	0	0
1992	0	1	1	0	0	0	0	0	10	0	0	0	3	0	5	0	0
1993	0	1	1	0	0	0	0	0	11	0	0	0	4	0	3	0	0
1994	0	1	1	0	0	0	0	0	11	0	0	0	5	0	2	0	0
1995	0	1	1	0	0	0	0	0	10	0	0	0	4	0	2	0	0
1996	0	1	1	0	0	0	0	0	11	0	0	0	4	0	2	0	0
1997	0	1	1	0	0	0	0	0	10	0	0	0	4	0	2	0	0
1998	0	1	1	0	0	0	0	0	10	0	0	0	3	0	2	0	0
1999	0	1	1	0	0	0	0	0	11	0	0	0	3	0	0	0	0
2000	0	1	1	0	0	0	0	0	10	0	0	0	3	0	0	0	0
2001	0	1	0	0	0	0	0	0	9	0	0	0	1	0	0	0	0
2002	0	1	0	0	0	0	0	0	7	0	0	0	1	0	0	0	0

We regressed measured fine mass concentrations against b_{sp} (for $b_{sp} < 800$ Mm^{-1}) and used the regression coefficients to predict fine mass from b_{sp}. We then identified outliers, which we defined as points for which our predictions of fine mass exceeded measured levels of PM$_{10}$ or TSP mass concentrations. We excluded the suspect outliers and repeated the regressions of fine mass against the b_{sp} measurements.

For years prior to 1995, the nephelometer data correlated well ($r^2 > 0.8$) with the fine PM measurements from the dichotomous samplers (Figure F1). However, the nephelometer measurements were poorly correlated with both dichot and FRM fine PM mass concentrations from 1995 to 2002 (Figure F1). We were unable to determine the cause of the difference and recommend that further investigation be carried out; for our purposes, the earlier data were of more value (later years have reasonably extensive measurements of fine PM from dichotomous or FRM samplers). Although site-specific regressions in the later years exhibited better agreement than did regressions with all sites included (Figure F1), one location showed an unexplained regression shift (Figure F1b) and site-specific regressions were not particularly useful for us: to improve our data coverage, we wished to compute fine PM mass concentrations from nephelometer measurements at 15 to 20 sites (Table F1) using generic regression coefficients determined from the five sites having collocated fine PM mass and nephelometer measurements.

We attempted to refine the predictions of fine mass from b_{sp} by incorporating measurements of maximum and daily-average RH and temperature, as well as precipitation. No obvious improvements were obtained in the correlation coefficients. Therefore, we incorporated the b_{sp} measurements that were made prior to 1995 for predicting fine PM mass concentrations, and excluded all nephelometer data from 1995 to the present. Following the procedure used for special-study data (Appendix D), we predicted the dichot fine PM mass concentrations using a no-intercept regression as:

F1.   Predicted dichot fine PM mass = $0.14 * b_{sp}$ (+/- 0.0013)

The predictions of dichot fine PM mass for 1988-94 reproduced the measured dichot fine mass concentrations well ($r^2 > 0.8$) (Figure F2), whereas predictions for later years did not.
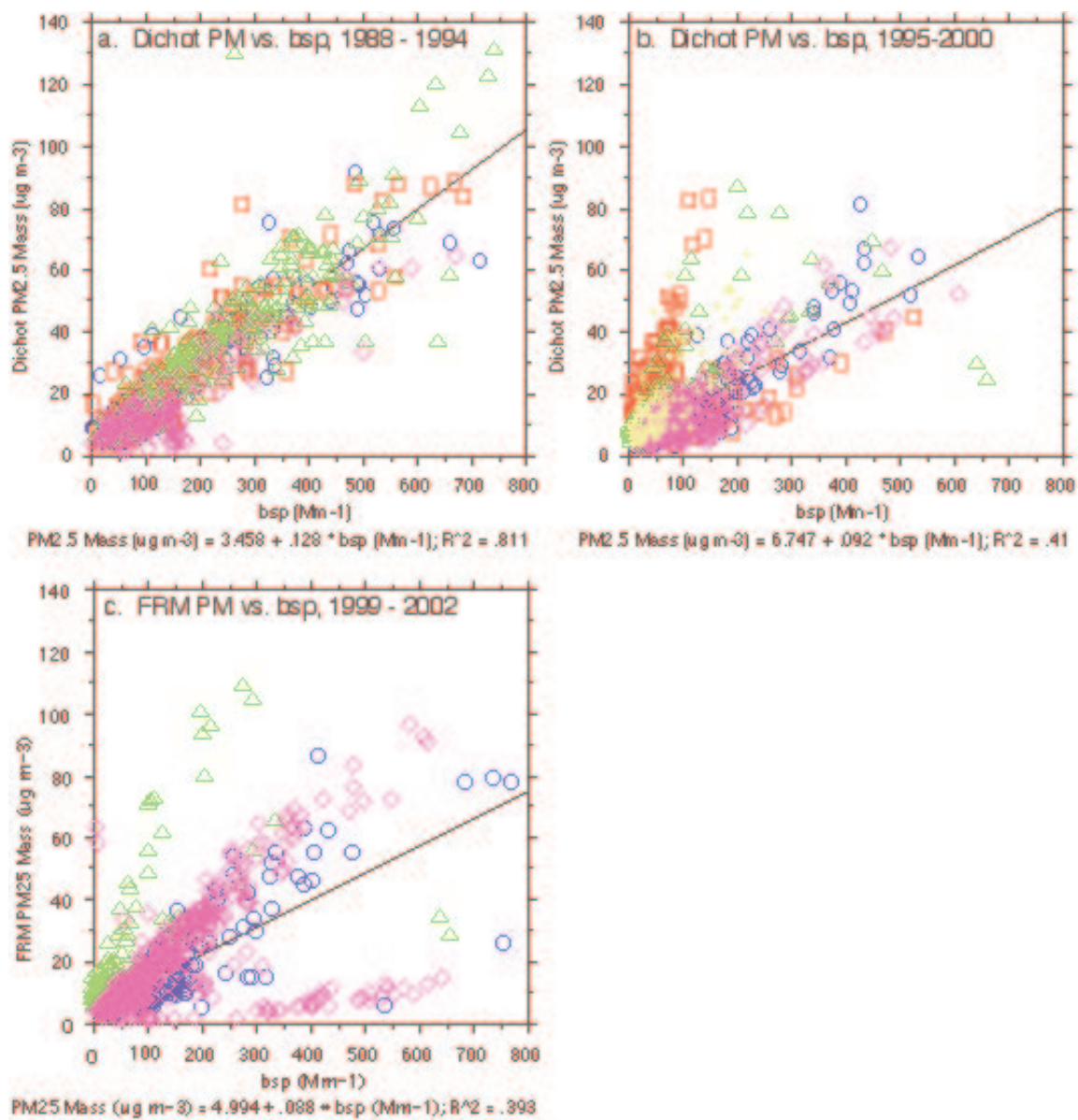
Figure F1. Fine PM mass concentrations from dichotomous samplers versus light scattering (b_{sp}). No data were reported from dichotomous samplers after year 2000.
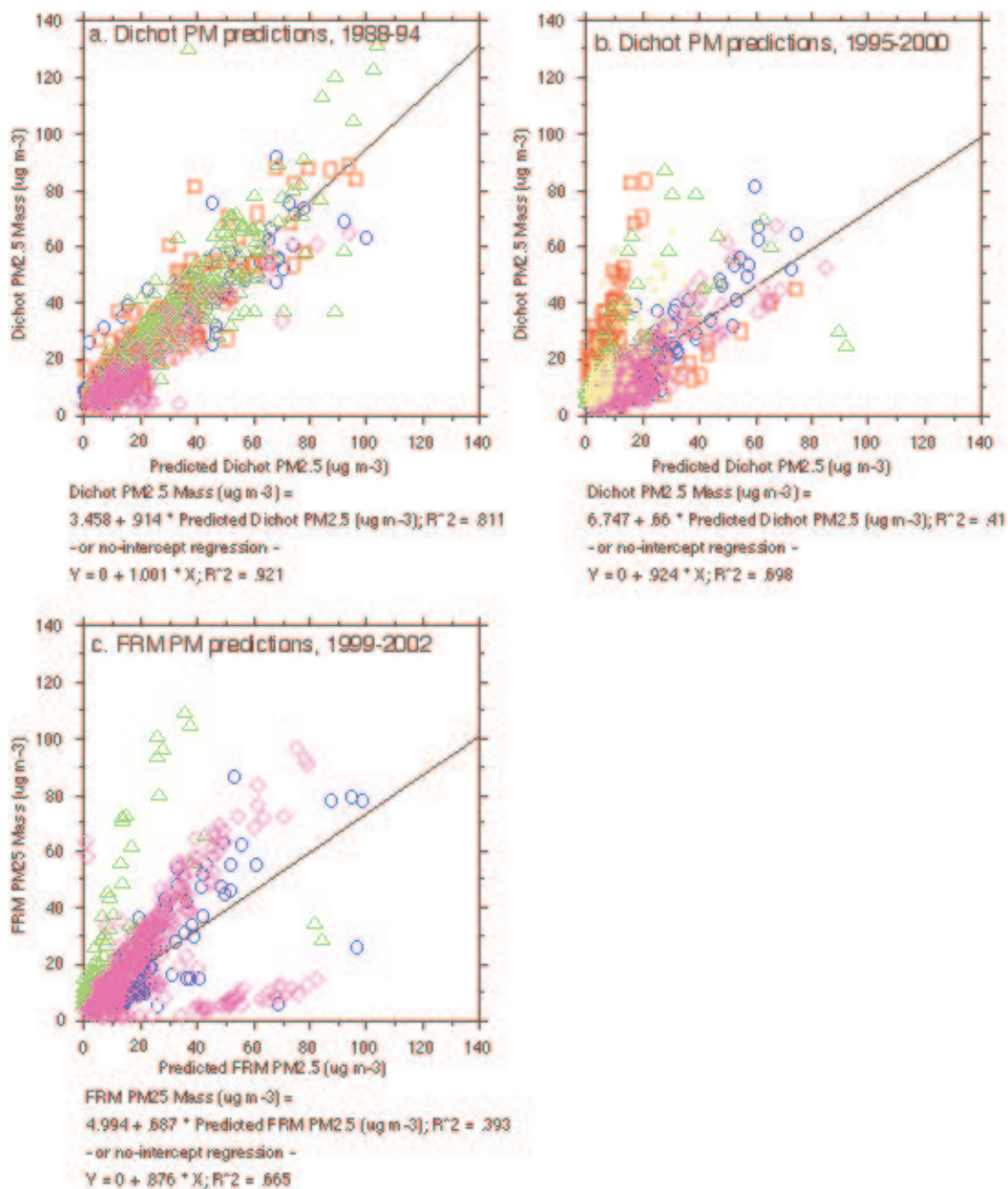
Figure F2. Fine PM mass concentrations from dichotomous samplers versus predictions made from light scattering (b_{sp}) measurements. No data were reported from dichotomous samplers after year 2000.

The final estimation step was conversion of the dichot-equivalent to FRM-equivalent fine mass concentrations, as described in Appendix C. We reconstructed fine PM mass concentrations for all the sites listed in Table F1, subject to the procedures and exclusions documented above.

APPENDIX G. ERROR ANALYSIS

In Appendix B, we characterized the uncertainties of the monthly averages. It is also of interest, but more difficult, to characterize the estimation errors, i.e., the differences between predicted and true monthly averages. If the true monthly averages were known, of course, there would be no need to reconstruct estimates from other measurements. However, it is possible to use the FRM measurements as a standard of comparison for the period of time when the FRM data are available. In this appendix, we summarize the frequency of measurement types and examine the differences between FRM and other monthly averages for evidence of bias, or systematic error. We compare the magnitudes of systematic error to our computed uncertainties, and determine the degree of intersite correlation among the errors.

Measurement Frequency

The complete database consists of best-estimate monthly averages for each of the sites listed in Appendix A. Table G1 lists the total number of site-months by predominant measurement type.

Table G1. Numbers of site-months with best-estimate monthly-average fine PM mass, by measurement type. For any month that included days having fine PM mass measurements or reconstructions from multiple methods, the month's measurement type was categorized as the method used for the greatest number of sampling days during the month. Of the total, 2274 site month had fewer than four sampling days.

Source of Best Estimate PM2.5	Site-months
Total	14545
Nephelometer	2871
FRM	3610
Special studies PM2.5	774
Dichot PM2.5, site-specific	707
Dichot PM2.5, generic	1074
Reconstructed from PM10 SO4 & NO3 plus carbon from CoH, site-specific	860
Reconstructed from PM10 SO4 & NO3 plus carbon from CoH, generic	1260
Reconstructed from PM10 SO4 & NO3 plus carbon from CO, site-specific	293
Reconstructed from PM10 SO4 & NO3 plus carbon from CO, generic	2553
Reconstructed from PM10 SO4, NO3 & carbon, generic	543

Seasonality

We compared the daily-average measurements of fine PM mass from dichot samplers, converted to FRM-equivalent units as discussed in Appendix C, to data from FRM samplers and computed the differences. Similarly, we computed daily-average reconstructed fine PM mass from measurements of sulfate, nitrate, and carbon, converted to FRM-equivalent units as discussed in Appendix E, to data from FRM samplers and computed the differences. The distributions of the differences are shown in Figure G1. The median dichot-FRM differences are essentially zero during all months, indicating that no seasonal bias exists, although the range of errors is larger during winter months, especially November through February, than in other months. The median differences between reconstructed and FRM fine mass concentrations vary somewhat more from month to month, but the variation of the medians is small (less than $\sim$ 1 to 3 μg m^{-3}) and is not systematic. As with the dichot samplers, the range of differences between reconstructed and FRM fine mass is greater during winter months. Since mean fine PM concentrations are greater during winter months than at other times, it is possible for the differences to be greater during winter. No systematic seasonal bias exists for any of the estimates of FRM fine mass. Note, however, that because larger differences tend to occur during winter months, it is possible that such differences (or prediction errors) could show some degree of correlation among monitoring locations.
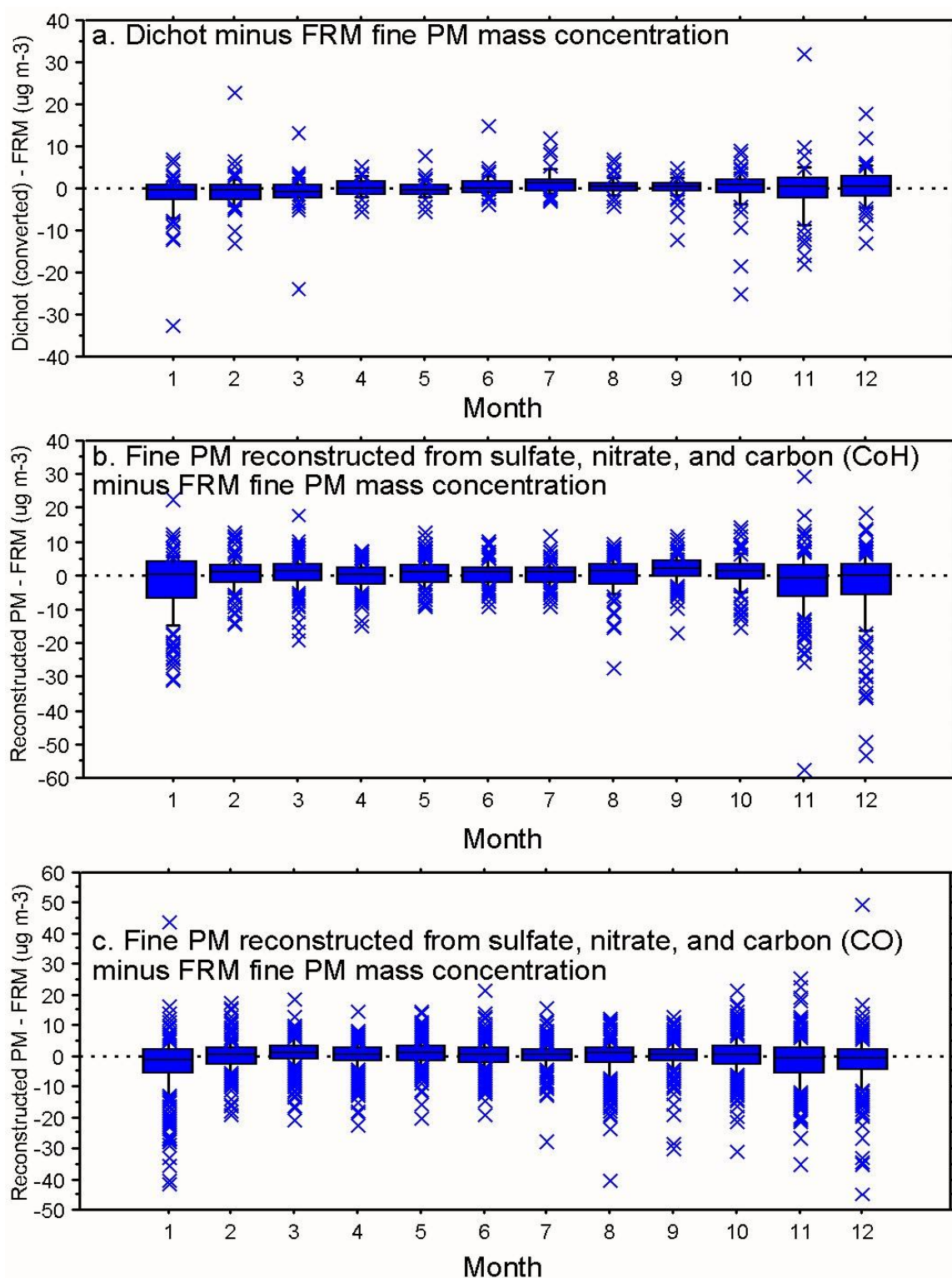
Figure G1. Distributions of differences between measured or reconstructed fine mass concentrations and FRM fine mass concentrations versus month. The box-and-whiskers plots denote the 10th, 25th, 50th, and 90th percentiles.

Trend

The differences between dichot or reconstructed fine PM mass concentrations and FRM fine mass concentrations exhibited some tendency to covary among sites and to vary over time (Figure G2). However, the magnitudes of the temporal variations were small and not readily amenable to correction, because the length of overlapping records was short and the trends, or drift, cannot be assumed to project backward in a linear fashion. In the case of the dichot and FRM samplers, some differences may exist in calibration schedules; such differences might or might not have occurred during earlier time periods. In the case of the reconstructed fine PM concentrations, either differences in calibration schedules or changes in PM composition over time could contribute to the drift. We note that for the period shown, the best estimates would be FRM measurements wherever and whenever they were available. The drift in the differences betweens reconstructed and FRM measurements was smaller than for the dichot drift, and opposite in direction.

Comparisons of reconstructed fine PM mass with dichot fine PM mass showed minimal drift over the 14-year span of overlapping monitoring record (Figure G3). Again, the magnitudes of the errors varied with season, even though the mean errors did not.
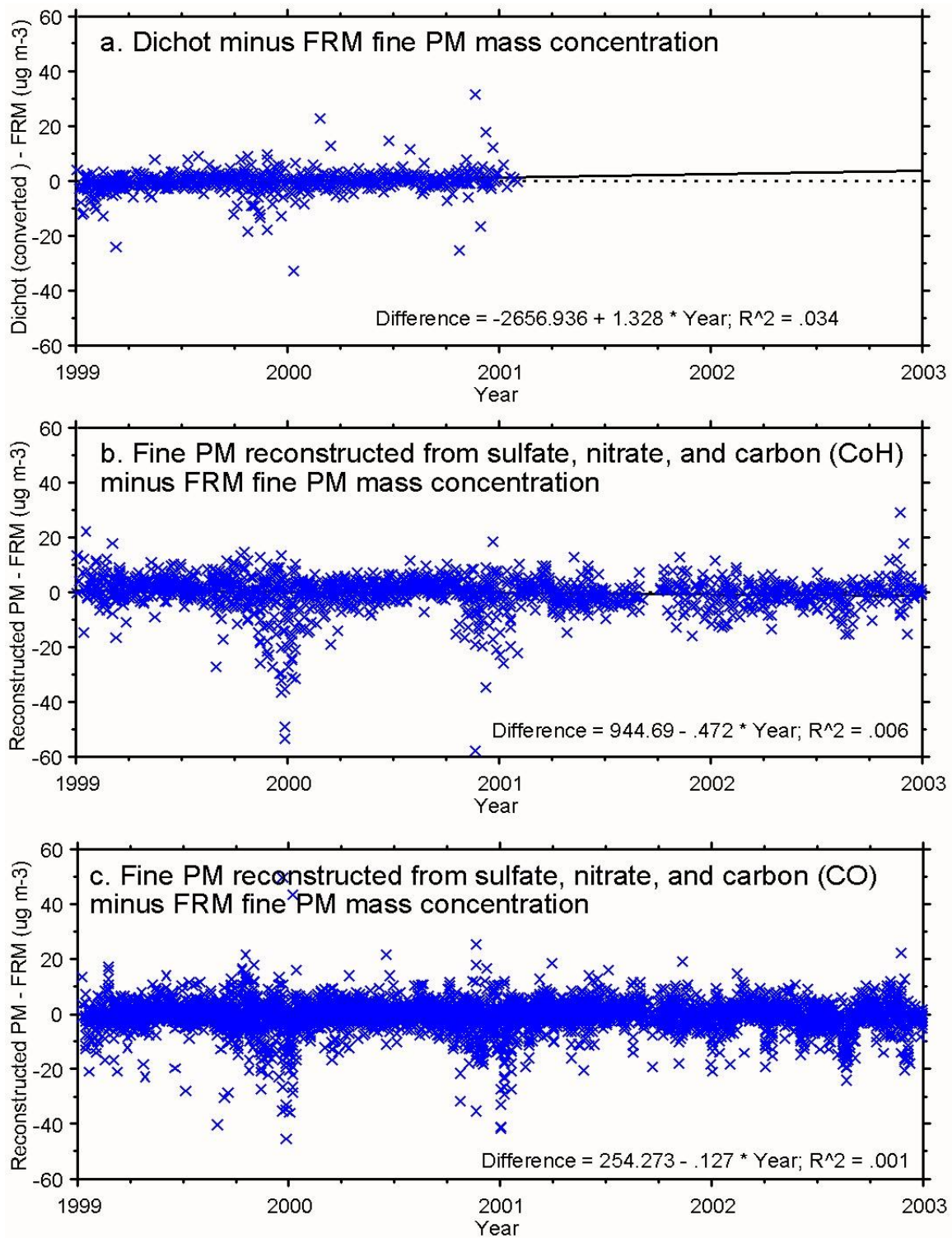
Figure G2. Differences between daily-average measured or reconstructed fine mass concentrations and FRM fine mass concentrations versus time.
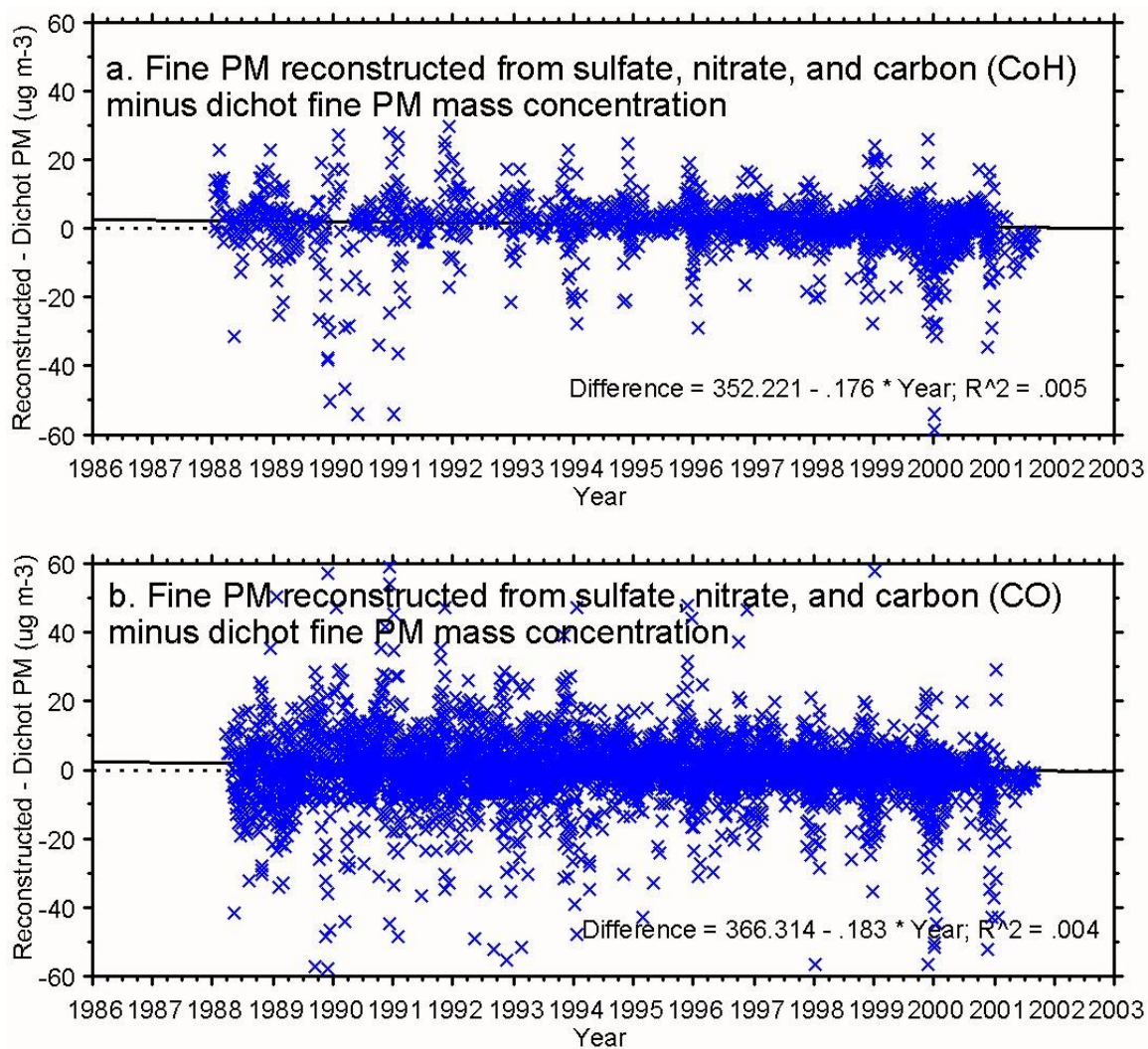
Figure G3. Differences between daily-average reconstructed fine mass concentrations and dichot fine mass concentrations versus time.

Intersite Correlation of Errors

As noted above, the differences between daily-average PM predictions and FRM measurements varied by only marginal amounts by season or over time. However, the magnitudes of the differences were greatest during some months, typically, November through February. As a result, some intersite covariance of the daily-average differences occurred (Figure G4). The correlations are a function of distance and fall off over approximately 200 km.

The presence of correlated errors in the daily-average reconstructed fine PM mass concentrations potentially leads to intersite correlation of errors in the monthly-average best estimates of fine PM mass concentrations, if the best estimates for different sites tend to be based upon the same measurement methods during the same time periods. This situation tends to occur, of course, though not necessarily at all sites. That is, the best estimates for the period 1999 through 2002 largely derive from FRM samplers at all locations, while the estimates for the earliest years (e.g., 1980 through 1985) tend to derive from the nephelometer data (Table G1). Between 1986 and 1998, the best estimates for different sites derived from dichot samplers, special studies, PM_{10} sulfate and nitrate in combination with estimates of carbon, or nephelometer measurements in the priority order documented in the Section II of this report, depending upon the availability of data at any particular site.
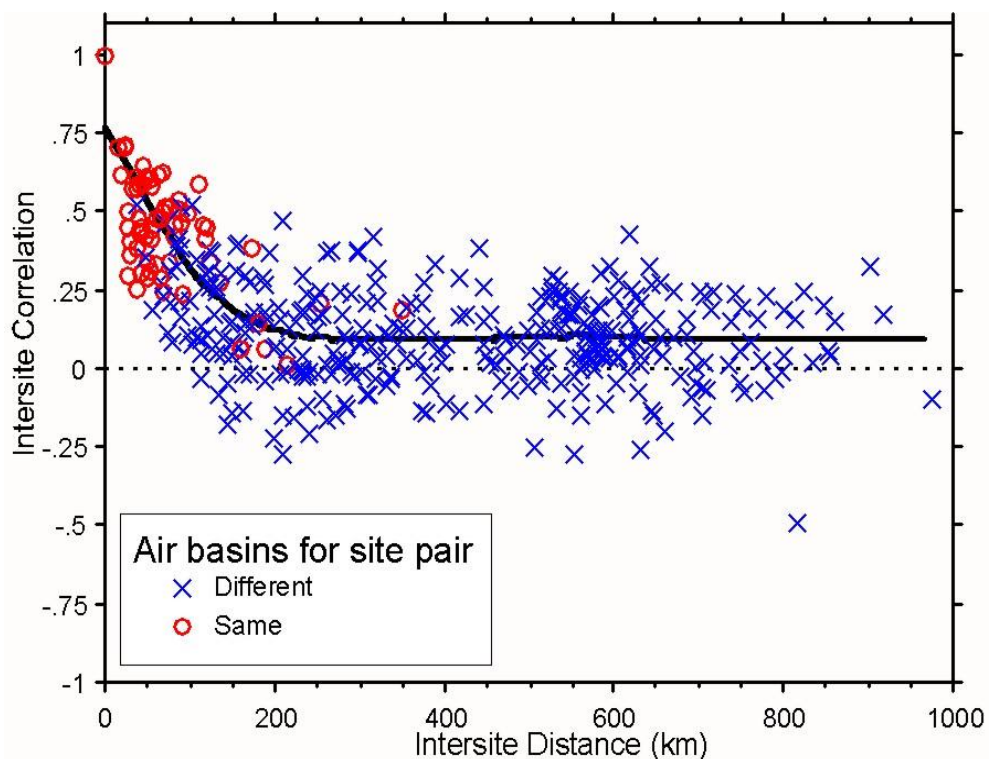
Figure G4. Intersite correlation of prediction errors versus intersite distance. The prediction errors were computed as the differences between the daily-average fine PM mass concentrations that were reconstructed from the sum of PM_{10} sulfate, PM_{10} nitrate, and carbon from CO, minus the fine PM mass concentration measured by collocated FRM samplers. The data are from 29 samplers throughout California and are shown in Figure G2c.

Table G1. Number of site-months of monthly-average fine PM mass concentration, disaggregated by year and measurement source. The columns are arranged (left to right) in the order of priority for choosing the best estimate of fine PM mass (except that monthly averages of the two-week sampler measurements were used as best estimates only for locations having no daily measurements).

YEAR	Measured Fine PM Mass				Reconstructed Fine PM Mass			
	FRM	Dichot	Special Studies	Two-Week Sampler	PM_{10} sulfate, nitrate, & carbon	PM_{10} sulfate, nitrate, & carbon from CoH	PM_{10} sulfate, nitrate, & carbon from CO	Light scattering (nephe-lometer)
1980								83
1981								109
1982			118					164
1983								203
1984						17	14	207
1985						67	85	206
1986			60			162	115	220
1987						189	169	204
1988		77	31			134	131	213
1989		127	70			129	170	210
1990		125	74		15	145	164	219
1991		154	65		43	140	157	198
1992		156	72		46	132	158	217
1993		162	82		40	137	182	209
1994		154	69	62	34	145	182	209
1995		186	91	59	25	188	168	
1996		195	17	78	130	139	220	
1997		193	11	73	159	138	221	
1998	4	196	14	87	37	119	311	
1999	810	38		60	14	46	177	
2000	914	10		55		48	86	
2001	968	8		59		26	64	
2002	914			67		19	72	

We compared monthly averages from dichot samplers to monthly averages from FRM samplers and computed the difference. We then checked the level of intersite correlation in the time series of differences. This assessment was limited by the number of sites having overlapping dichot and FRM measurements. We used sites having at least 20 months of collocated monthly averages, of which there were six: Stockton Hazelton St, Modesto 14th St, Fresno First St, Sacramento T St, Imperial Valley East Belcher St, and Bakersfield California Ave. Each monthly average included at least 4 sampling days, but the FRM samplers typically operated more days per month (averaging 14.8 days for the FRM and 5.9 days for the dichot samplers). Twelve of the 15 intersite correlations were not statistically significant (Table G2). The results do not reveal the presence of substantial spatial correlation of the errors.

Table G2. Spearman correlation coefficients of the differences between monthly-average dichot and FRM sampler fine mass concentrations. These intersite correlations were determined from 17 to 23 months of measurements, varying among site pairs. Non-redundant correlation coefficients that are statistically significant at the $p<0.05$ level are shown in bold type. When one month (December 2000) was excluded, only the Stockton-Fresno correlation was significant ($p<0.05$).

	Stockton	Modesto	Fresno	Sacramento	Imperial	Bakersfield
Stockton	1.00	0.11	**0.68**	**0.49**	0.19	0.09
Modesto	0.11	1.00	-0.01	0.19	**0.45**	0.22
Fresno	0.68	-0.01	1.00	0.36	0.36	-0.08
Sacramento	0.49	0.19	0.36	1.00	0.16	-0.08
Imperial	0.19	0.45	0.36	0.16	1.00	-0.22
Bakersfield	0.09	0.22	-0.08	-0.08	-0.22	1.00

To further examine the intersite correlation pattern, if any, we compared monthly averages from reconstructions of fine PM mass (from PM_{10} sulfate and nitrate plus carbon estimated from CoH) to monthly averages from the FRM samplers and computed the difference. We again checked the level of intersite correlation in the time series of differences, and found that 10 of the 36 intersite correlations were statistically significant (Table G3). These intersite correlations are again related to the occurrence or absence of larger differences during certain months (e.g., December 1999) at multiple sites. Such differences, in turn, might occur either because of the tendency for large errors to correlate, as previously noted for daily-average prediction errors, or because the reconstructed monthly averages were based on five sampling days, whereas the FRM averages ranged from 10 to 23 days. The difference in sampling frequency is a potential second source of bias in monthly averages.

Table G3. Spearman correlation coefficients of the differences between monthly-average fine mass reconstructed from PM_{10} sulfate, PM_{10} nitrate, and coefficient of haze minus monthly-average FRM sampler fine mass concentrations. These intersite correlations were determined from 12 to 25 months of measurements, varying among site pairs. Non-redundant correlation coefficients that are statistically significant at the $p<0.05$ level are shown in bold type.

	Visalia	Santa Rosa	Chico	Redwood City	Concord	Modesto	Fresno	Sacramento	Bakersfield
Visalia	1.00	**0.67**	**0.48**	0.38	0.22	**0.54**	**0.70**	**0.43**	**0.62**
Santa Rosa	0.67	1.00	**0.57**	0.46	0.00	**0.54**	0.48	0.39	**0.52**
Chico	0.48	0.57	1.00	-0.08	0.21	0.44	0.57	0.51	0.41
Redwood City	0.38	0.46	-0.08	1.00	0.02	-0.13	-0.15	0.11	0.11
Concord	0.22	0.00	0.21	0.02	1.00	0.15	0.31	0.11	0.22
Modesto	0.54	0.54	0.44	-0.13	0.15	1.00	0.18	0.23	**0.57**
Fresno	0.70	0.48	0.57	-0.15	0.31	0.18	1.00	0.37	0.45
Sacramento	0.43	0.39	0.51	0.11	0.11	0.23	0.37	1.00	0.13
Bakersfield	0.62	0.52	0.41	0.11	0.22	0.57	0.45	0.13	1.00

We also compared monthly averages from reconstructions of fine PM mass (from PM_{10} sulfate and nitrate plus carbon estimated from CO) to monthly averages from the FRM samplers and computed the difference. We again checked the level of intersite correlation in the time series of differences, and found that 45 of the 91 intersite correlations were statistically significant (Table G4). As in Table G3, some significant correlations are more closely related to the occurrence or absence of larger differences during certain months at multiple sites (again, December 1999; also, January 2001). The reconstructed monthly averages were based on 4 to 11 sampling days (mean 5.1), whereas the FRM averages ranged from 4 to 31 days (mean 12.7).

Since the largest intersite correlations for the monthly-average prediction errors were observed for the reconstructions of fine PM mass from PM_{10} sulfate and nitrate plus carbon estimated from CO, we plotted them as a function of intersite distance (Figure G5). The intersite correlations were largest for site pairs located in the same air basins (usually, the San Joaquin Valley), and fell off less rapidly with distance than did the intersite correlations of the daily-average prediction errors (compare Figure G4). The larger intersite correlations for monthly averages than for daily averages supports the previously-noted point that sampling frequency may contribute to prediction errors in the monthly averages. The estimation of monthly averages from 4 to 5 sampling days can be problematic, especially during winter when the value obtained on a particular day is strongly affected by the occurrence of specific weather conditions. In the San Joaquin Valley, for example, winter PM levels are dramatically different during multiday stagnation episodes compared with days when frontal systems pass through.

Table G4. Spearman correlation coefficients of the differences between monthly-average fine mass reconstructed from PM_{10} sulfate, PM_{10} nitrate, and CO minus monthly-average FRM sampler fine mass concentrations. These intersite correlations were determined from 13 to 31 months of measurements, varying among site pairs. Non-redundant correlation coefficients that are statistically significant at the p<0.05 level are shown in bold type.

	Visalia	Stockton	Chico	San Bernardino	Fremont	Vallejo	Riverside	Modesto	Roseville	El Rio	Fresno	Sacramento	Clovis	Bakersfield
Visalia	1.00	**0.80**	**0.48**	-0.10	0.40	0.41	0.13	**0.75**	0.36	**0.48**	**0.73**	**0.65**	**0.89**	**0.70**
Stockton	0.80	1.00	**0.51**	0.02	0.38	0.29	0.01	**0.76**	**0.49**	**0.45**	**0.82**	**0.75**	**0.78**	**0.57**
Chico	0.48	0.51	1.00	-0.12	0.43	**0.63**	0.16	**0.58**	**0.69**	**0.45**	0.31	**0.75**	0.34	**0.43**
San Bernardino	-0.10	0.02	-0.12	1.00	-0.04	0.01	**0.58**	0.01	-0.04	0.37	-0.16	0.27	-0.31	0.04
Fremont	0.40	0.38	0.43	-0.04	1.00	**0.68**	0.00	0.29	**0.66**	0.15	0.46	**0.56**	0.46	0.30
Vallejo	0.41	0.29	0.63	0.01	0.68	1.00	0.24	0.42	**0.53**	0.23	0.37	**0.61**	0.37	0.28
Riverside	0.13	0.01	0.16	0.58	0.00	0.24	1.00	0.01	0.07	0.35	0.06	0.28	0.04	0.12
Modesto	0.75	0.76	0.58	0.01	0.29	0.42	0.01	1.00	**0.58**	**0.55**	**0.71**	**0.58**	**0.64**	**0.71**
Roseville	0.36	0.49	0.69	-0.04	0.66	0.53	0.07	0.58	1.00	0.18	**0.43**	**0.68**	0.29	**0.37**
El Rio	0.48	0.45	0.45	0.37	0.15	0.23	0.35	0.55	0.18	1.00	**0.53**	0.39	0.34	**0.49**
Fresno	0.73	0.82	0.31	-0.16	0.46	0.37	0.06	0.71	0.43	0.53	1.00	**0.69**	**0.68**	**0.68**
Sacramento	0.65	0.75	0.75	0.27	0.56	0.61	0.28	0.58	0.68	0.39	0.69	1.00	**0.58**	**0.55**
Clovis	0.89	0.78	0.34	-0.31	0.46	0.37	0.04	0.64	0.29	0.34	0.68	0.58	1.00	**0.56**
Bakersfield	0.70	0.57	0.43	0.04	0.30	0.28	0.12	0.71	0.37	0.49	0.68	0.55	0.56	1.00
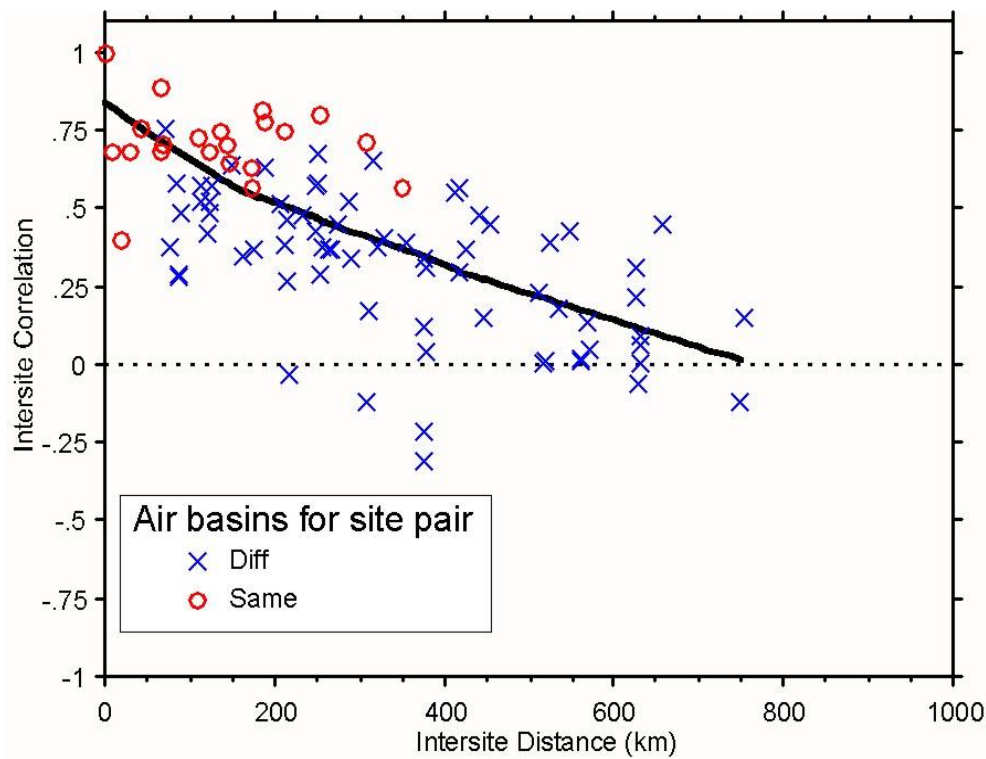
Figure G5. Intersite correlation of prediction errors versus intersite distance. The prediction errors were computed as the differences between the monthly-average fine PM mass concentrations that were reconstructed from the sum of PM_{10} sulfate, PM_{10} nitrate, and carbon from CO, minus the monthly-average fine PM mass concentration measured by collocated FRM samplers. The data are from 11 samplers throughout California as listed in Table G3.

The degree to which intersite correlations of prediction errors result in correlation of errors in the best estimates of monthly average PM concentrations depends upon the numbers and locations of sites for which the best estimates were based upon the same types of reconstructed PM concentrations (Table G1). At many sites, the best estimates included days having a variety of measurement types. For example, some months at one location might include five days of dichot PM measurements and 25 days of PM mass estimated from nephelometer data. At other locations, the best estimates during some periods might consist of the same 5 sampling days with reconstructed fine PM mass. This information is included within the database. Tables G5 and G6 list the sites, by year, where one or more monthly-average best estimates of PM fine mass concentration consist of reconstructions from PM_{10} species and estimates of carbon concentrations, and where fewer than six sampling days were available. Users of the database may find this information helpful for identifying time periods and site pairs whose distance and measurement type might result in correlated estimation errors.

Table G5. List of sites, by year, where one or more monthly-average best estimates of PM fine mass concentration were reconstructed from PM_{10} species and estimates of carbon concentrations from CoH, and where fewer than six sampling days per month were available.

Site Name	'84	'85	'86	'87	'88	'89	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
Bakersfield-5558 California Ave	x	x	x	x															
Bethel Island Road			x	x	x	x	x	x											
Calexico-Ethel Street											x	x	x						
Chico-Manzanita Avenue												x			x				
Citrus Heights-Sunrise Blvd					x														
Concord-2975 Treat Blvd			x	x	x	x	x	x	x	x	x	x	x	x	x	x			
El Cajon-Redwood Avenue								x	x	x	x	x	x	x					
Fremont-Chapel Way											x	x	x	x	x				
Fresno-Olive Street	x	x	x	x			x												
Livermore-Old 1st Street			x	x	x	x	x	x	x	x	x	x							
Mammoth Lakes-Gateway HC		x	x	x	x		x												
Napa-Jefferson Avenue			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Oceanside-Mission Avenue								x	x	x	x	x	x	x					
Oildale-3311 Manor Street			x							x		x	x	x	x	x	x	x	x
Paso Robles-Santa Fe Ave											x								
Quincy-N Church Street						x	x												
Quincy-S Redburg Avenue			x	x															
Redwood City			x	x	x	x	x	x	x	x	x	x	x	x	x	x			
Richmond-13th Street						x	x	x	x	x	x	x							
Rocklin-Sierra College			x	x	x														
Roseville-N Sunrise Blvd										x		x	x	x	x				
Salinas-Natividad Road #2		x	x	x	x														
San Francisco-Arkansas Street			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
San Jose-4th Street	x	x	x	x	x	x													
San Jose-W San Carlos Street						x	x	x	x	x	x	x							
San Luis Obispo-Marsh Street				x	x	x	x	x	x	x	x	x	x	x			x		
San Rafael			x	x	x	x	x	x	x	x	x	x							
Santa Maria-906 S Broadway																	x		
Santa Rosa-5th Street												x	x	x	x	x			
Simi Valley-Cochran I		x																	
Simi Valley-Cochran Street		x	x	x	x	x	x	x											
South Lake Tahoe-3377 Tahoe Blvd					x														
Stockton-Hazelton Street		x																	
Vallejo-304 Tuolumne Street												x	x						
Visalia-N Church Street	x	x	x	x															
Willits-Firehouse		x																	
Willows-E Laurel Street												x	x	x	x	x	x	x	
Yosemite Village-Visitor Center			x	x															

Table G6. List of sites, by year, where one or more monthly-average best estimates of PM fine mass concentration were reconstructed from PM_{10} species and estimates of carbon concentrations from CO, and where fewer than six sampling days per month were available.

Site Name	'84	'85	'86	'87	'88	'89	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
Anaheim-Harbor Blvd							x	x	x	x	x	x	x	x	x				
Avalon-Crescent Avenue							x												
Azusa				x								x	x			x			
Barstow		x	x	x	x	x	x	x	x	x	x	x	x						
Bethel Island Road									x	x	x	x	x	x	x	x	x	x	x
Burbank-W Palm Avenue	x	x		x		x			x	x	x	x	x	x	x		x		
Calexico-Ethel Street												x	x						
Chula Vista			x	x	x														
Clovis-N Villa Avenue									x	x	x	x	x	x	x		x		
Concord-2975 Treat Blvd					x														
El Cajon-Redwood Avenue			x	x	x	x	x	x						x	x				
El Centro-9th Street														x	x	x			
El Rio-Rio Mesa School					x	x	x	x											
El Rio-Rio Mesa School #2					x	x	x	x	x	x	x	x	x	x	x	x			
El Toro	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
Fontana-Arrow Highway		x	x	x	x	x	x	x											
Fresno-Cal State #2		x	x	x	x	x													
Fresno-Olive Street			x	x															
Goleta			x	x	x														
Hawthorne							x	x	x	x	x	x	x	x	x	x	x	x	x
Hesperia-Olive Street							x		x	x	x	x	x	x					
Lancaster						x	x												
Lancaster-W Pondera Street							x	x	x	x	x	x	x	x					
Livermore-Old 1st Street												x	x	x	x	x			
Los Angeles-North Main Street		x		x	x		x	x	x		x		x	x					
Mammoth Lakes-Gateway HC									x		x	x							
Mexicali-CBTIS														x	x	x			
Mexicali-Cobach														x	x	x			
Mexicali-ITM														x	x	x			
Mexicali-UABC														x	x	x			
Napa-Jefferson Avenue															x				
North Long Beach	x	x		x															
Oceanside-Mission Avenue	x	x	x	x	x	x	x	x						x	x				
Palm Springs-Fire Station				x	x	x	x	x	x	x	x	x	x	x	x	x			
Pittsburg-10th Street																x	x	x	x
Redding-Health Dept Roof											x								
Richmond-13th Street									x			x	x	x					
Riverside-Rubidoux	x	x		x									x	x	x				
Rosarito														x	x	x			
San Bernardino-4th Street			x	x	x	x	x				x	x	x	x	x				
San Diego-Logan Avenue																x	x	x	

Site Name	'84	'85	'86	'87	'88	'89	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02
San Diego-Overland Avenue			x																
San Francisco-Arkansas Street												x	x						
San Jose-W San Carlos Street											x	x							
San Luis Obispo-Marsh Street							x											x	
San Pablo-El Portal														x	x				x
San Pablo-Rumrill Blvd																			x
San Rafael								x				x	x	x	x	x	x	x	x
Santa Clarita																		x	x
Santa Clarita-County Fire Station						x	x	x	x	x	x	x	x	x	x	x	x	x	
Santa Clarita-Honby						x	x												
Santa Maria-906 S Broadway																	x	x	
Simi Valley-Cochran I		x																	
Simi Valley-Cochran Street		x	x		x	x	x	x	x	x	x	x	x	x	x				
South Lake Tahoe-3377 Tahoe Blvd		x	x	x	x														
Tecate-Paseo Morelos																x			
Temecula-Rancho California Road								x	x	x									
Tijuana-ITT												x		x	x				
Tijuana-La Mesa															x	x			
Tijuana-Las Playas														x	x	x			
Twentynine Palms-Adobe Road										x									
Twentynine Palms-Adobe Road #2											x	x	x						
Vallejo-304 Tuolumne Street												x	x	x	x	x			
Victorville-Armagosa Road										x	x	x							
Visalia-N Church Street				x															
Weaverville-Hospital				x															
Willits-Firehouse			x																
Yreka-Foothill Drive				x	x														

Comparison of Estimated Uncertainties With Prediction Errors

For each monthly average best-estimate of fine PM mass concentration in the database, we report an estimated uncertainty (1 and 2 sigma). Because the uncertainties are calculated quantities (see Appendix B), we would like to know how well they represent the true estimation errors. Here, we compare the distributions of the uncertainties to the differences between monthly-average PM estimates and monthly-average FRM fine mass concentrations (Figure G6). If the reported uncertainties represent true uncertainty reasonably well, we would expect that approximately 95 percent of the differences would be within the 2 sigma uncertainty limits. This result holds. The results show that approximately 95 percent of the calculated uncertainties were less than 20 μg m^{-3}, varying among site months (Figure G6). About the same percentage of differences between dichot or reconstructed fine mass and FRM fine mass concentrations were within +/- 20 μg m^{-3}.
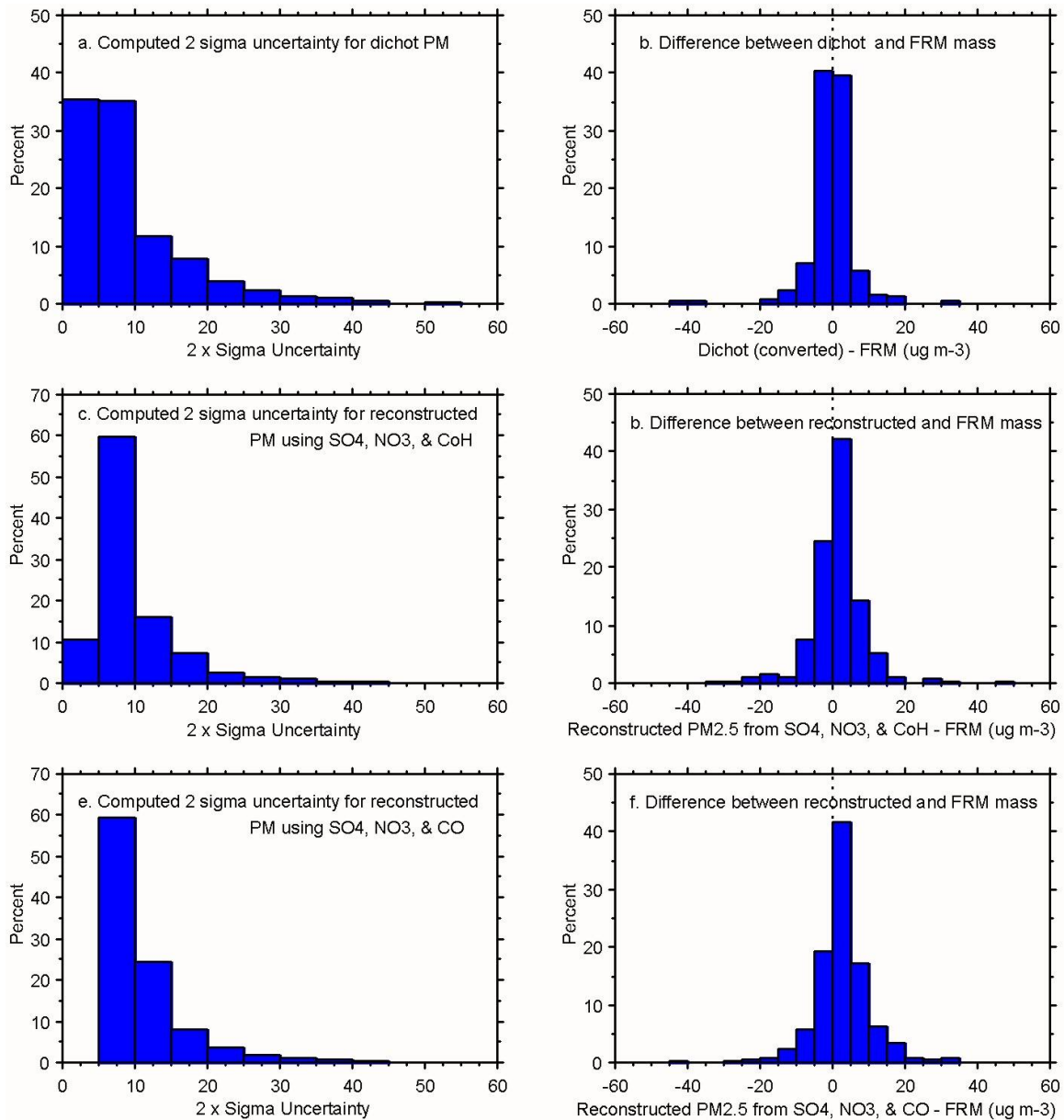
Figure G6. Comparison of estimated uncertainties with prediction errors. Panels a, c, and e show the distributions of estimated uncertainties for best-estimate monthly fine PM averages consisting of (a) dichot mass concentrations, (b) PM reconstructed from sulfate, nitrate, and CoH, and (c) PM reconstructed from sulfate, nitrate, and CO. Panels b, d, and f show the differences between monthly-average dichot or reconstructed fine PM mass and FRM fine PM mass concentrations.

APPENDIX H. COMPUTER PROGRAMS

<u>Directory of Programs</u>

I. Data Programs
  a. Data8002.sas
  b. Select.sas
  c. DayCOall.sas
  d. Adjustal.sas
  e. SpstDat4.sas

II. Regression Programs
  a. RegdYRpm.sas
  b. RegCOsit.sas
  c. RegPMal2.sas

III. Monthly Average Program
  a. EachMn4g.sas

```
**************************************************************************
*********;
*data8002.sas;

*Program combines data from individual years and converts to ppbv from
ppmv;
*The CO data will be used;

********** Input files **************
dat1980.sd2 through dat2002.sd2 were based on
ARB files HRO31980.dat through HRO32002.dat
***********************************;

*********** Output file ************
data8002.sd2
********************************;

*Define SAS library;
libname sasfiles 'c:\work\arb data';
*combine raw data from 1980-2002;
data temp;
set sasfiles.dat1980 sasfiles.dat1981 sasfiles.dat1982 sasfiles.dat1983
sasfiles.dat1984
    sasfiles.dat1985 sasfiles.dat1986 sasfiles.dat1987 sasfiles.dat1988
sasfiles.dat1989
    sasfiles.dat1990 sasfiles.dat1991 sasfiles.dat1992 sasfiles.dat1993
sasfiles.dat1994
    sasfiles.dat1995 sasfiles.dat1996 sasfiles.dat1997 sasfiles.dat1998
sasfiles.dat1999
    sasfiles.dat2000 sasfiles.dat2001 sasfiles.dat2002 ;
*convert from ppmv to ppbv;
o31=o3*1000;
no1=no*1000;
nox1=nox&1000;
no21=no2*1000;
co1=co*1000;
nmhc1=nmhc*1000;

drop o3 no nox no2 co nmhc;

data sasfiles.data8002  ;
set temp;
*rename variables;
o3=o31;
no=no1;
no2=no21;
nox=nox1;
nmhc=nmhc1;
co=co1;
drop o31 no1 no21 nox1 nmhc1 co1;

run;

************************** End of Program **********************;
```

```
***********************************************************************
***********;
*select.sas;


*Program combines data sets from dlypm files and pm2510 files -- uses
only one monitor
  for any given day from pm2510 dataset;
*Also selects 4 sites (azusa, elmonte,stockton/hazleton and sacto T st)
with full data record
 for testing;

*********Input data ***************
PM2510.sd2 (combination of pm25daily.txt and pm10stddaily.txt files
from ARB CD)
dlypm8.sd2  ARB file
dlypm9.sd2  ARB file
*********************************;

**********Output data ************
PM8002.sd2   ARB PM data for all available sites
select.sd2   ARB PM data for Azusa, El Monte, Stockton/Hazelton and
Sacramento/T St
********************************;


*Define SAS library;
libname sasfiles 'e:\work\arb-pm';

data temp1;
set sasfiles.pm2510;
keep site monitor month day year pm10nat pm25nat basin;
proc sort;
  by site year month day monitor;
proc sort nodupkey;  *eliminate all but one monitor value for each site
and date;
 by site year month day;

data temp2;
set sasfiles.dlypm8 sasfiles.dlypm9;
if(loc_code le '9999');  *keep only sites,not summary data;
site=loc_code +1-1;      *convert to numeric site codes;
proc sort;
  by site year month day;

*combine pm10, pm25 and speciation;
data sasfiles.PM8002;
merge  temp1 temp2;
  by site year month day;
*add in basins when missing;
if(basin eq 'SC') then basin_lt='X';
if(basin eq 'SCC') then basin_lt='W';
if(basin eq 'NC') then basin_lt='M';
if(basin eq 'NCC') then basin_lt='L';
if(basin eq 'GBV') then basin_lt='C';
```

```sas
if(basin eq 'LC') then basin_lt='F';
if(basin eq 'LT') then basin_lt='G';
if(basin eq 'MEX') then basin_lt='5';
if(basin eq 'MD') then basin_lt='I';
if(basin eq 'MC') then basin_lt='J';
if(basin eq 'NEP') then basin_lt='N';
if(basin eq 'SV') then basin_lt='Q';
if(basin eq 'SS') then basin_lt='R';
if(basin eq 'SD') then basin_lt='S';
if(basin eq 'SFB') then basin_lt='T';
if(basin eq 'SJV') then basin_lt='U';
IF(basin eq 'OUT') then basin_lt='6';

*select sites;

data sasfiles.select;
set sasfiles.pm8002;
if(site eq 2484 or site eq 2813 or site eq 2094 or site eq 3011);
*azusa 2484, el monte 2813, stockton/hazleton 2094, sacramento t st
3011;

run;

***************************** End of Program
*************************;
```

```
***********************************************************************
*;
*dayCOall;

*************** Input Files ***************
data8002.sd2 created from data8002.sas
****************************************;

*************** Output File *************
co8002al.sd2  daily average CO
***************************************;

*Define SAS library;
libname sasfile2 'c:\work\arb data';

*get co hourly data;
data temp;
set sasfile2.data8002;

if(co ne .);

proc sort;
  by loc_code year month day;

*average daily CO;
proc means noprint;
 var co;
 by loc_code year month day;
 output out=tempco
 mean=co
 n=nco;

data sasfile2.co8002al;
set tempco;
if(nco ge 18);  *make sure there are at least 18 hr in each day of
data;

run;

***************************** End of Program
***********************;
```

```
****************************************************************;
*Adjustal.sas;


*Program calculates adjustments for elevation, to be used with dichot
measurements
  in later programs;  *this version does so for all sites;

*Define SAS library;
libname sasfiles 'c:\work\arb-pm';


*note that several sites have missing elevations, recorded as "0";
*Input file is location.dat, downloaded from
www.arb.ca.gov/aqd/aqdcd/aqdcddld.htm;
*when elevation is 0, adjust = 1;

data sasfiles.adjustal;
set sasfiles.location;
if(site ne 0);      *keep only site info, not basin info;
adjust=exp(-0.1146*elev/1000);
 keep site adjust elev;
proc sort;
  by site;

*Output file records adjustment factor as variable, adjust;

run;

****************************End of
Program****************************************;
```

```
***********************************************************************
**********;

* SPSTDAT4.SAS;

********* Calculate predictions for pm25 from special study data for
pm25 *******;


**************** Define SAS libraries ****************;

libname sasfile2 'c:\work1\arb\specialstudydata\cadmp';
libname sasfile3 'c:\work1\arb\specialstudydata\ptep';
libname sasfile4 'c:\work1\arb\specialstudydata\vaqs';
libname sasfile5 'c:\work1\arb\specialstudydata\caltech';
libname sasfiles 'c:\work\arb-pm';

*****************************************************;

* Output consists of PM2.5 predictions, recorded in spstdat4.sd2;
* Revised July 6, 2005;

**************** Input files **********************;
*
CADMP    c:\work1\arb\specialstudydata\cadmp\pm25tf.sd2
PTEP     c:\work1\arb\specialstudydata\ptep\ana25.sd2  (Anaheim)
                                          dbr25.sd2  (Diamond Bar)
                                          dla25.sd2  (downtown L.A.)
                                          ftn25.sd2  (Fontana)
                                          rub25.sd2  (Rubidoux)
VAQS     c:\work1\arb\specialstudydata\vaqs\vaqs8889.sd2
CALTECH c:\work1\arb\specialstudydata\caltech\pmfine.sd2
                                          pmfine82.sd2

pmfine93.sd2
c:\work\arb-pm\ADJUSTAL.SD2     Adjustment factors based on elevation to
                                convert from STP to ambient conditions;


************************************************************;


*Access CADMP data and assign standard CARB location codes;
data temp1;
set sasfile2.pm25tf;
if(location eq 'Azusa') then site=2484;
if(location eq 'Bakersfi') then site=3146;
if(location eq 'North Lo') then site=2429;
if(location eq 'Sacramen') then site=3011;
if(location eq 'Freemont') then site =2293;
if(location eq 'Gasquet') then site=3027;
if(location eq 'LA North') then site=2899;
if(location eq 'Santa Ba') then site=2708;  *Goleta;
if(location eq 'Sequoia-') then site =2069; *giant forest;
if(location eq 'Yosemite') then site=3018;  *turtleback dome;
```

```sas
*there is overlap between caltech data and cadmp data for 1993, sites
2484,2899 and
  2429.  There is more caltech data, so use caltech instead of cadmp
for
  1993 and these sites only;
if(year eq 1993 and (site eq 2899 or site eq 2484 or site eq 2429) )
then pm25tf=.;
proc sort nodupkey;
  by site year month day;

* Add elevation adjustment factors and calculate PM2.5 predictions
***********;
data temp11;
merge temp1 sasfiles.adjustal;
  by site;
if(pm25tf ne .);
pm25=pm25tf;
preddi=pm25tf*0.914;
s2predDI=( (1+(1/318))*(8260.459/316) ) + ( ((15.063-pm25tf)**2) *
(0.01427**2) );
predFRM=-0.017 + (1.16*preddi*adjust);
s2predFR=( (1+(1/1053))*(14842.754/1051) ) + ( ((16.661-
(preddi*adjust))**2) * (0.007439**2) );

*Access VAQS data and assign standard CARB location codes;
data temp2;
set sasfiles.vaqs8889;
if(instr eq 'VAQSA');
year=year+1900;  *correct for year coding;
location=site;
drop site;

data temp222;
set temp2;
if(location eq 'CLD') then site=2437;   *crows landing-davis;
if(location eq 'COV') then site=2638;   *corcoran-van dorsten;
if(location eq 'FPT') then site=2617;   *five points;
if(location eq 'KCC') then site=2916;   *kettleman city- cal trans;
if(location eq 'LAV') then site=2972;   *lassen volcanic;
if(location eq 'FOV') then site=2367;   *fresno-Olive St;
if(location eq 'FEL') then site=3024;   *Taft College;
if(location eq 'KRW') then site=2181;   *kern refulge;
if(location eq 'SOH') then site=2094;   *stockton hazelton;
*exclude VAQS data for BAK  because there is CADMP data for the same
years;
proc sort;
  by site;

* Add elevation adjustment factors and calculate PM2.5 predictions
***********;
data temp22;
merge temp222 sasfiles.adjustal;
  by site;
if(result ne .);
pm25=result;
preddi=pm25*0.716;
```

```
s2predDI=( (1+(1/100))*(7623.169/98) ) + ( ((34.705-pm25)**2) *
(0.02**2) );
predFRM=-0.017 + (1.16*preddi*adjust);
s2predFR=( (1+(1/1053))*(14842.754/1051) ) + ( ((16.661-
(preddi*adjust))**2) * (0.007439**2) );
keep location year month day pm25 preddi s2preddi predfrm s2predfr;


*Access PTEP data and asign standard CARB location codes;
data temp3;
set sasfile3.ana25 ;
location='ANA';
pm25=tm;
keep location year month day pm25;
data temp4;
set sasfile3.dbr25 ;
location='DBR';
pm25=tm;
keep location year month day pm25;
data temp5;
set sasfile3.dla25 ;
location='DLA';
pm25=tm;
keep location year month day pm25;
data temp6;
set sasfile3.ftn25 ;
location='FTN';
pm25=tm;
keep location year month day pm25;
data temp7;
set sasfile3.rub25 ;
location='RUB';
pm25=tm;
keep location year month day pm25;

data temp8;
set temp3 temp4 temp5 temp6 temp7;
if(location eq 'ANA') then site=2623;
if(location eq 'DBR') then site=3130;
if(location eq 'DLA') then site=2899;
if(location eq 'FTN') then site=2266;
if(location eq 'RUB') then site=2596;
proc sort;
  by site;

* Add elevation adjustment factors and calculate PM2.5 predictions
***********;
data temp88;
merge temp8 sasfiles.adjustal;
  by site;
 if(pm25 eq -9) then pm25=.;  *missing values coded as "-9";
if(pm25 ne .);
preddi=pm25*0.845;
s2predDI=( (1+(1/41))*(490.993/39) ) + ( ((29.733-pm25)**2) *
(0.01599**2) );
predFRM=-0.017 + (1.16*preddi*adjust);
```

```sas
s2predFR=( (1+(1/1053))*(14842.754/1051) ) + ( ((16.661-
(preddi*adjust))**2) * (0.007439**2) );

*Access CALTECH data and asign standard CARB location codes;
data tempcal;
set sasfile5.pmfine sasfile5.pmfine82 sasfile5.pmfine93;
if(mass lt 0) then mass=.;
if(sta eq 60) then site=2484;  *azusa;
if(sta eq 72) then site=2429;  *n long beach;
if(sta eq 87) then site=2899;  *la n main;
if(sta eq 144) then site=2596; *rubidoux;
if(sta eq 200) then site=3672; *san nicolas is;
if(sta eq 175) then site=2485; *upland;
if(sta eq 176) then site=2623; *anaheim;
if(sta eq 69) then site=2492; *burbank;
if(sta eq 76) then site=2045; *hawthorne;
if(sta eq 400) then site=. ; *claremont--no regular arb site;
if(sta eq 86) then site=2494; *w la;
if(sta eq 83) then site=2160; *pasadena;
if(sta eq 100) then site=.; *not on site list;
if(sta eq 300) then site=.; *tanbark flats--no regular arb site;

proc sort;
  by site;
* Add elevation adjustment factors and calculate PM2.5 predictions
***********;
data tempcal2;
merge tempcal sasfiles.adjustal;
  by site;
if(mass ne .);
pm25=mass;
preddi=mass*0.898;
s2predDI=( (1+(1/159))*(7642.745/157) ) + ( ((26.593-mass)**2) *
(.01709**2) );
predFRM=-0.017 + (1.16*preddi*adjust);
s2predFR=( (1+(1/1053))*(14842.754/1051) ) + ( ((16.661-
(preddi*adjust))**2) * (0.007439**2) );


*combine special studies;
data SASFILES.spstDat4;
set temp11 temp22 temp88 tempcal2;
s2pred=s2predfr + s2predDI;

if(site ne .);
*exclude vaqs data for BAK  because there is cadmp data for the same
years;

run;

*************** END OF PROGRAM
*****************************************;
```

```
**************************************************************************
**********;
*regdYRpm.sas;

*********** Program computes linear regression coefficients, site
specific *******;


*define SAS library;
libname sasfiles 'c:\work\arb-pm';

*************** ;

* Output consists of regression coefficients, recorded in PM25parc.sd2;

*************** Input files ************************;

* PM8002.SD2     PM and light scattering 24-hour data from CARB for
1980 - 2002;

**************************************************;

*************** Variable definitions ***************;


* PM25NAT      FRM PM2.5 mass;
* PMFINE       Dichot PM2.5 mass;

* CO           carbon monoxide;
* cohav24      24-hour average coefficient of haze (CoH);
* LTSCAT       light scattering (units of 10**-4 meters);
* bsp          light scattering due to particles (Mm-1);

* pms4n3       pm10s04 + pm10no3;
* pms4n3TC     pm10s04 + pm10no3 +pm10totC;
* pms4n3C      pm10s04 + pm10no3 + C estimated from CoH;
* pms4n3CO     pm10s04 + pm10no3 +(co*.008);
* tsps4n3      tspso4 + tspno3;
* tsps4n3C     tspso4 + tspno3 + C estimated from CoH ;


****************************************************;

* Regressions are site-specific;

* Adjustments are made for elevation when using dichot pm25 and SSI
pm10 measurements
  in generic regressions to account for conversion from STP to ambient
(FRM);

* Adjustments for elevation are not made in site-specific regressions;


****************************************************;
```

```
****************** Access PM data file **************;


data tempdat;
set sasfiles.pm8002;
****************** Remove suspect data and correct below-detects
*************************;

if(site eq 3146 and year eq 2001 and (month ge 1 and month le 3) ) then
cohav24=.;
if(cohav24 eq 0) then cohav24=.001;  *it doesn't make sense to say
cohav24 is zero,
                                     so put in a very low number
(lower than any recorded);
C=((3.4*71)/10) * (cohav24**0.76) ;

******** Run QA checks on estimated carbon concentrations
****************;

*Remove C if C from pm10totC or C calculated from CoH or from CO is gt
 PM10 mass minus measured inorganic components because
 carbon should not exceed leftover mass (pmMASS-sulfate,nitrate and
ammonium);
pmnh4=pm10nh4;   *calculate pm10nh4 when necessary;
if(pm10nh4 eq .) then do;
  pmnh4=(18/2) * ((pm10s04/48) + (pm10no3/62) );
  end;
pm=pm25nat - pm10s04 - pm10no3 - pmnh4;
if(pm25nat eq .) then pm=pmfine - pm10s04 - pm10no3 - pmnh4;

if(c gt pm and pm ne .) then c=.;
if(pm10totc gt pm and pm ne .) then pm10totc=.;

if(pm25nat eq . and pmfine eq .) then do;
   pm10=pm10nat - pm10s04 - pm10no3 - pmnh4;
    if(c gt pm10 and pm10 ne .) then c=.;
    if(pm10totc gt pm10 and pm10 ne .) then pm10totc=.;
   end;

************ Define predictor variables without correction for STP to
ambient ********;

pms4n3=pm10s04 + pm10no3;
pms4n3TC=pm10s04 + pm10no3 +pm10totC;

pms4n3C=pm10s04 + pm10no3 + C;
tsps4n3=tspso4 + tspno3;
tsps4n3C=tspso4 + tspno3 + C;


**************** Calculate regression coefficients
*******************;

%macro mm;
data temp;
set tempdat;
```

```
y=&yvar;
x=&xvar;

xy=x*y;
x2=x**2;
y2=y**2;

if(x ne . and y ne .);
SPECIES=&spec;
INDEPENT=&XVARname;
DEPENDNT=&YVARname;
proc sort;
  by species INDEPENT dependnt &fin2;

proc means noprint;
var x y xy x2 y2;
by  species INDEPENT dependnt &fin2;

output out=temp2
       mean=mx my mxy mx2 my2
       n=nx;

data &fout;
set temp2;

Sxx=nx*(mx2 - (mx**2));
Syy=nx*(my2 - (my**2));
Sxy=nx*(mxy - (mx*my));

if(nx gt 1);
slope=Sxy/Sxx;
intercep=my-(slope*mx);

SSR=slope*Sxy;
SSE=Syy-(slope*Sxy);

if(nx gt 2) then s2=SSE/(nx-2);

r2=SSR/Syy;

seslope=sqrt(s2/Sxx);
seinter=sqrt(s2*mx2/Sxx);

siglev=2*(1-probt(abs(slope/seslope),(nx-2)));


 run;
%mend;

 %let yvar=pm25nat; %let xvar=pmfine; %let fin2=site;  %let
spec='1Y=PM25NAT X=PMFINEDI';
   %let xvarname='PMFINEDI';  %let yvarname='PM25NAT ';     %let
fout=PMFINERC; %mm;
  %let yvar=pm25nat; %let xvar=PMS4N3TC; %let fin2=site;  %let
spec='2Y=PM25NAT X=PMS4N3TC';
```

```
      %let xvarname='PMS4N3TC';    %let yvarname='PM25NAT ';     %let
fout=PMSNTCRC; %mm;
 %let yvar=pm25nat; %let xvar=PMS4N3C; %let fin2=site;  %let
spec='3Y=PM25NAT X=PMS4N3C';
      %let xvarname='PMS4N3C ';   %let yvarname='PM25NAT ';     %let
fout=PMSNCRC; %mm;
 %let yvar=pm25nat; %let xvar=PMS4N3; %let fin2=site;  %let
spec='4Y=PM25NAT X=PMS4N3';
       %let xvarname='PMS4N3  ';   %let yvarname='PM25NAT ';     %let
fout=PMS4N3RC; %mm;
 %let yvar=pm25nat; %let xvar=TSPSO4; %let fin2=site;  %let
spec='5Y=PM25NAT  X=TSPSO4';
        %let xvarname='TSPSO4  ';   %let yvarname='PM25NAT ';   %let
fout=PMSO4RC; %mm;
 %let yvar=pmfine; %let xvar=PMS4N3TC; %let fin2=site;  %let
spec='6Y=PMFINE  X=PMS4N3TC';
        %let xvarname='PMS4N3TC';   %let yvarname='PMFINE  ';   %let
fout=fPMSNTCR; %mm;
 %let yvar=pmfine; %let xvar=PMS4N3C; %let fin2=site;  %let
spec='7Y=PMFINE  X=PMS4N3C';
        %let xvarname='PMS4N3C ';     %let yvarname='PMFINE  ';  %let
fout=FPMSNCR; %mm;
 %let yvar=pmfine; %let xvar=PMS4N3; %let fin2=site;  %let
spec='8Y=PMFINE  X=PMS4N3';
         %let xvarname='PMS4N3  ';   %let yvarname='PMFINE  ';   %let
fout=FPMS4N3R; %mm;
%let yvar=pmfine; %let xvar=TSPSO4; %let fin2=site;  %let
spec='9Y=PMFINE   X=TSPSO4';
        %let xvarname='TSPSO4  ';     %let yvarname='PMFINE  ';   %let
fout=FPMSO4RC; %mm;




*COMBINE REGRESSION RESULTS and CREATE PERMANENT DATASET FOR OUTPUT;
data sasfiles.pm25PArc;
set pmfinerc pmsntcrc pmsncrc pms4n3rc PMSO4RC  FPMSNTCR FPMSNCR
FPMS4N3R FPMSO4RC;

run;

************************* End of Program *************************;
```

```
**************************************************************************
****;
*regCOsit;
* Program computes linear regression coefficients to predict pm25nat
from
  PM10SO4 + PM10NO3 + co for each site;
*not log-transform;

******************** Input Files ***************
co8002al.sd2  Daily CO averaged from ARB hourly data files
pm8002.sd2    PM and light scattering 24-hour data from CARB for 1980 -
2002;
**********************************************;

****************** Output File *************
PMcoSIrc.sd2
**********************************************;

*************** Variable definitions ***************;


* PM25NAT     FRM PM2.5 mass;
* CO          carbon monoxide;
* pms4n3CO    pm10s04 + pm10no3 +(co*.008);


****************************************************;

*define SAS libraries;
libname sasfile2 'c:\work\arb data';
libname sasfiles 'c:\work\arb-pm';



*get co daily data ;
data temp;
set sasfile2.co8002al;
proc sort;
  by loc_code year month day;

*combine daily co with pm measurements;
data temp3;
set sasfiles.pm8002;
proc sort;
  by loc_code year month day;

data tempdat;
merge temp3 temp;
  by loc_code year month day;


*define predictor variable;
pms4n3CO=pm10s04 + pm10no3 + (co*.008);

%macro mm;
```

```
data temp;
set tempdat;


y=&yvar;
x=&xvar;

xy=x*y;
x2=x**2;
y2=y**2;

if(x ne . and y ne .);
SPECIES=&spec;
proc sort;
  by species &fin2;

proc means noprint;
var x y xy x2 y2;
by  species &fin2;

output out=temp2
       mean=mx my mxy mx2 my2
       n=nx;

data &fout;
set temp2;

Sxx=nx*(mx2 - (mx**2));
Syy=nx*(my2 - (my**2));
Sxy=nx*(mxy - (mx*my));

if(nx gt 1);
slope=Sxy/Sxx;
intercep=my-(slope*mx);

SSR=slope*Sxy;
SSE=Syy-(slope*Sxy);

if(nx gt 2) then s2=SSE/(nx-2);

r2=SSR/Syy;

seslope=sqrt(s2/Sxx);
seinter=sqrt(s2*mx2/Sxx);

siglev=2*(1-probt(abs(slope/seslope),(nx-2)));


 run;
%mend;

 %let yvar=pm25nat; %let xvar=pms4n3co; %let fin2=site;  %let
spec='PMS4N3CO';
          %let fout=PMSNCOrc; %mm;

*Create permanent dataset;
data sasfiles.pmCOsirc;
```

```
        set pmsncorc;

run;
                 ******************************** End of Program
                     *****************************;
```

```
**************************************************************************
**********;

* REGPMAL2.SAS;

*********** Program computes linear regression coefficients *******;


**************** Define SAS libraries ****************;
libname sasfile2 'c:\work\arb data';
libname sasfiles 'c:\work\arb-pm';

*************** ;

* Output consists of regression coefficients, recorded in PM25A2rc.sd2;
* Revised July 6, 2005;

**************** Input files ***********************;

* PM8002.SD2      PM and light scattering 24-hour data from CARB for
1980 - 2002;
* ADJUSTAL.SD2    Adjustment factors based on elevation to convert from
STP to ambient conditions;


**************************************************;

*************** Variable definitions ***************;


* PM25NAT     FRM PM2.5 mass;
* PMFINE      Dichot PM2.5 mass;

* CO          carbon monoxide;
* cohav24     24-hour average coefficient of haze (CoH);
* LTSCAT      light scattering (units of 10**-4 meters);
* bsp         light scattering due to particles (Mm-1);

* pms4n3      pm10s04 + pm10no3;
* pms4n3TC    pm10s04 + pm10no3 +pm10totC;
* pms4n3C     pm10s04 + pm10no3 + C estimated from CoH;
* pms4n3CO    pm10s04 + pm10no3 +(co*.008);
* tsps4n3     tspso4 + tspno3;
* tsps4n3C    tspso4 + tspno3 + C estimated from CoH;


**************************************************;

* Regressions are generic (not site-specific);

* Adjustments are made for elevation when using dichot pm25 and SSI
pm10 measurements
  in generic regressions to account for conversion from STP to ambient
(FRM);
```

111

```
******************************************************;

****************** Access PM data file **************;

data tempdat;
set sasfiles.pm8002;
*if(month ge 4 and month le 9);    *comment out to keep all dates;

****************** Identify relocated sites to be combined
*************;

*combine Bakersfield Chester with Bakersfield California -- transition
occurred in April 1994;
if(site eq 2131) then site=3146;

*************** Remove suspect data and correct below-detects
*****************;

if(site eq 3146 and year eq 2001 and (month ge 1 and month le 3) ) then
cohav24=.;
if(cohav24 eq 0) then cohav24=.001;  *it doesn't make sense to say
cohav24 is zero,
                                      so put in a very low number
(lower than any recorded);
C=((3.4*71)/10) * (cohav24**0.76) ;



proc sort;
  by site year month day;

****************** Add in CO data *********************************;
data tempco;
set sasfile2.co8002al;
site=loc_code +1-1;
proc sort;
  by site year month day;

data tempdat2;
merge tempdat tempco;
  by site year month day;



DATA TEMPDAT3;
SET TEMPdat2;

************ Add elevation adjustment factors to data ***********;

data tempadj;
set sasfiles.adjustal;
keep site adjust;
proc sort;
  by site;

data tempdat4;
merge tempdat3 tempadj;
```

112

```
  by site;

************* Define carbon estimated from CO ************;

CfromCO=0.008*CO;

******** Run QA checks on estimated carbon concentrations
****************;

*Remove C if C from pm10totC or C calculated from CoH or from CO is gt
 PM10 mass minus measured inorganic components because
 carbon should not exceed leftover mass (pmMASS-sulfate,nitrate and
ammonium);


pmnh4=pm10nh4;    *calculate pm10nh4 when necessary;
if(pm10nh4 eq .) then do;
  pmnh4=(18/2) * ((pm10s04/48) + (pm10no3/62) );
  end;
pm=pm25nat - pm10s04 - pm10no3 - pmnh4;
if(pm25nat eq .) then pm=pmfine - pm10s04 - pm10no3 - pmnh4;

if(c gt pm and pm ne .) then c=.;
if(pm10totc gt pm and pm ne .) then pm10totc=.;
if(cfromCO gt pm and pm ne .) then cfromCO=.;

if(pm25nat eq . and pmfine eq .) then do;
    pm10=pm10nat - pm10s04 - pm10no3 - pmnh4;
    if(c gt pm10 and pm10 ne .) then c=.;
    if(pm10totc gt pm10 and pm10 ne .) then pm10totc=.;
      if(cfromCO gt pm10 and pm10 ne .) then cfromCO=.;
   end;

********** Define predictor variables with correction for STP to
ambient ********;

pmfine2=pmfine*adjust;
pms4n3=(pm10s04 + pm10no3)*adjust;
pms4n3TC=(pm10s04 + pm10no3 +pm10totC) * adjust;
pms4n3C=((pm10s04 + pm10no3) * adjust) + C;
pms4n3CO=((pm10s04 + pm10no3) * adjust) +cfromCO;

* These variables will not be used;
*tsps4n3=tspso4 + tspno3;
*tsps4n3C=tspso4 + tspno3 + C;


*** Create permanent SAS data set *********;
data sasfiles.reggen;
set tempdat4;
if(pm25nat ne .);

**************** Calculate regression coefficients
*******************;

************* Do not use option to calculate site-specific coefficient
*********;
```

```
%macro mm;
data temp;
set tempdat4;


y=&yvar;
x=&xvar;

xy=x*y;
x2=x**2;
y2=y**2;

if(x ne . and y ne .);
SPECIES=&spec;
INDEPENT=&XVARname;
DEPENDNT=&YVARname;
proc sort;
  by species INDEPENT dependnt ;   * remove &fin2=site for this version;

proc means noprint;
var x y xy x2 y2;
by  species INDEPENT dependnt ;      * remove &fin2=site for this
version;

output out=temp2
       mean=mx my mxy mx2 my2
       n=nx;

data &fout;
set temp2;

Sxx=nx*(mx2 - (mx**2));
Syy=nx*(my2 - (my**2));
Sxy=nx*(mxy - (mx*my));

if(nx gt 1);
slope=Sxy/Sxx;
intercep=my-(slope*mx);

SSR=slope*Sxy;
SSE=Syy-(slope*Sxy);

if(nx gt 2) then s2=SSE/(nx-2);

r2=SSR/Syy;

seslope=sqrt(s2/Sxx);
seinter=sqrt(s2*mx2/Sxx);

siglev=2*(1-probt(abs(slope/seslope),(nx-2)));


 run;
%mend;
```

```
 %let yvar=pm25nat; %let xvar=pmfine2;* %let fin2=site;   %let
spec='1Y=PM25NAT X=PMFINEDI';
   %let xvarname='PMFINEDI';  %let yvarname='PM25NAT ';     %let
fout=PMFINERC; %mm;
  %let yvar=pm25nat; %let xvar=PMS4N3TC; *%let fin2=site;  %let
spec='2Y=PM25NAT X=PMS4N3TC';
    %let xvarname='PMS4N3TC';    %let yvarname='PM25NAT ';     %let
fout=PMSNTCRC; %mm;
 %let yvar=pm25nat; %let xvar=PMS4N3C; *%let fin2=site;  %let
spec='3Y=PM25NAT X=PMS4N3C';
    %let xvarname='PMS4N3C ';   %let yvarname='PM25NAT ';     %let
fout=PMSNCRC; %mm;
 %let yvar=pm25nat; %let xvar=PMS4N3; *%let fin2=site;  %let
spec='4Y=PM25NAT X=PMS4N3';
     %let xvarname='PMS4N3  ';   %let yvarname='PM25NAT ';     %let
fout=PMS4N3RC; %mm;
 %let yvar=pm25nat; %let xvar=PMS4N3co; *%let fin2=site;  %let
spec='4Y=PM25NAT X=PMS4N3CO';
     %let xvarname='PMS4N3CO';   %let yvarname='PM25NAT ';     %let
fout=PMSNCORC; %mm;



************** COMBINE REGRESSION RESULTS AND CREATE OUTPUT
************;
data sasfiles.pm25A2rc;
set pmfinerc pmsntcrc pmsncrc pms4n3rc  pmsncorc;

run;

************** END OF PROGRAM
******************************************;
```

```
***************************************************************************
***********;

* EACHMN4G.SAS;

*********** Program computes best estimate monthly average fine PM mass
*******;


***** Define SAS libraries **************;

libname sasfile2 'c:\work\arb data';
libname sasfiles 'c:\work\arb-pm';

******************************************;

* Program generates monthly averages of measurements and predictions;
* Output is recorded in EachMn4g.sd2;
* Revised from eachMn3g.sas 6/29/05 and 6/30/05;
* Revised July 6, 2005;
* Revised September 1, 2005;

**************** Input files ***********************;

* PM8002.SD2      PM and light scattering 24-hour data from CARB for
1980 - 2002;
* CO8002AL.SD2     CO data 1980 - 2002;
* SPSTDAT4.SD2    Data from special studies;
* ADJUSTAL.SD2    Adjustment factors based on elevation to convert from
STP to ambient conditions;
* PM25A2RC.SD2    Generic regression coefficients (rc) determined from
all sites together;
* PM25PARC.SD2     Site-specific regression coefficients for four
predictors;
* PM25SIRC.SD2    Site-specific regression coefficients for one
predictor;


* The regression coefficients (rc) are factors needed to convert from
  various predictor concentrations to FRM fine mass concentration;

******************************************************;

**************** Variable definitions **************;


* PM25NAT      FRM PM2.5 mass;
* PMFINE       Dichot PM2.5 mass;

* CO           carbon monoxide;
* cohav24      24-hour average coefficient of haze (CoH);
* LTSCAT       light scattering (units of 10**-4 meters);
* bsp          light scattering due to particles (Mm-1);
```

```
* pms4n3       pm10s04 + pm10no3;
* pms4n3TC     pm10s04 + pm10no3 +pm10totC;
* pms4n3C      pm10s04 + pm10no3 + C estimated from CoH;
* pms4n3CO     pm10s04 + pm10no3 +(co*.008);
* tsps4n3      tspso4 + tspno3;
* tsps4n3C     tspso4 + tspno3 + C estimated from CoH;


*********************************************************;


* Predict best estimate PM25 using regression coefficients and
available predictors.
  Predictors are taken in the following preferred order, when the FRM
measurement
  of PM25 was unavailable:
     pmfine (dichot) from site specific rc,
     pmfine from generic rc,
     pm25 from special studies,
     pmso4no3totC with site specific rc,
     pmso4no3totC with generic rc,
     pms4n3coh with site specific rc,
     pms4n3coh with generic rc,
     pms4n3C-CO with site specific rc,
     pms4n3C-CO with generic rc,
     ltscat with generic rc,
     NOT pms4n3
     NOT tsps4n3
     NOT tsps4n3C;

* The monthly TWS data are added for comparison at the end;

* Adjustments are made for elevation when using dichot pm25 and SSI
pm10 measurements
  in generic regressions to account for conversion from STP to ambient
(FRM);


*********************************************************;

****************** Access PM data file **************;

data tempdat;
set sasfiles.pm8002;

******************** Identify relocated sites to be combined
**************;

*combine Bakersfield Chester with Bakersfield California -- transition
occurred in April 1994;
if(site eq 2131) then site=3146;

********** Remove suspect data and correct below-detects
********************;

if(year ge 1995)then ltscav24=.;     *do not use light scattering
measurements after 1994
```

```
*******************;

if(site eq 3146 and year eq 2001 and (month ge 1 and month le 3) ) then
cohav24=.;


if(cohav24 eq 0) then cohav24=.001;  *it doesn't make physical sense to
say cohav24 is zero,
                                      so put in a very low number
(lower than any recorded);

*********** Define carbon calculated from coefficient of haze (CARB
formula) ****;
C=((3.4*71)/10) * (cohav24**0.76) ;

*******************************************;

proc sort;
   by site year month day;

****** Add in CO data ********************;
data tempco;
set sasfile2.co8002al;
site=loc_code +1-1;

proc sort;
   by site year month day;

*******************************************;

data tempdat2;
merge tempdat tempco;
   by site year month day;

******** Run QA checks on estimated carbon concentrations
***************;

*Remove C if C from pm10totC or C calculated from CoH or from CO is gt
 PM10 mass minus measured inorganic components because
 carbon should not exceed leftover mass (pmMASS-sulfate,nitrate and
ammonium);

* Estimate C from CO;
C_co=co*.008;

     pmnh4=pm10nh4;   *calculate pm10nh4 when necessary;
     if(pm10nh4 eq .) then do;
       pmnh4=(18/2) * ((pm10s04/48) + (pm10no3/62) );
        end;

     *Use PM25 when available, otherwise use PM10 for this QA analysis;
     if(pm25nat ne . or pmfine ne .) then do;
       pm=pm25nat - pm10s04 - pm10no3 - pmnh4;
       if(pm25nat eq .) then pm=pmfine - pm10s04 - pm10no3 - pmnh4;

       if(c gt pm and pm ne .) then c=.;
       if(pm10totc gt pm and pm ne .) then pm10totc=.;
```

118

```sas
        if(c_co gt pm and pm ne .) then co=.;
    end;

    if(pm25nat eq . and pmfine eq . and pm10nat ne .) then do;
        pm10=pm10nat - pm10s04 - pm10no3 - pmnh4;
        if(c gt pm10 and pm10 ne .) then c=.;
        if(pm10totc gt pm10 and pm10 ne .) then pm10totc=.;
            if(c_co gt pm10 and pm10 ne .) then co=.;
        end;

    *********************************************************;


***************** Special Studies Data *****************;
*add in special studies data from cadmp,ptep,vaqs and caltech;
* SPSTDAT4 eliminates duplicate (collocated) CADMP data;
data tempsp;
set sasfiles.spstdat4;
pm25spst=pm25;
s2predsp=s2pred;
keep site year month day pm25spst s2predsp;
proc sort;
  by site year month day;

data tempMRG;
merge tempdat2 tempsp;
  by site year month day;

*********************************************************;

************* Define predictor variables *************;

DATA TEMPDAT3;
SET TEMPMRG;
data=1;

************* Define predictors without correction for STP to ambient
*************;
pms4n3=pm10s04 + pm10no3;
pms4n3TC=pm10s04 + pm10no3 +pm10totC;
pms4n3C=pm10s04 + pm10no3 + C;
tsps4n3=tspso4 + tspno3;
tsps4n3C=tspso4 + tspno3 + C;
pms4n3CO=pm10s04 + pm10no3 +(co*.008);

*********************************************************;

********* Define predictor variables with correction for STP to ambient
********;

data tempdat4;
merge tempdat3 sasfiles.adjustal;
  by site;
if(data eq 1);
pmfine2=pmfine*adjust;
pmsnTCg=((pm10s04 + pm10no3)*adjust) +pm10totC;
pms4n3Cg=((pm10s04 + pm10no3)*adjust) + C;
```

```
pmsnCOg=((pm10s04 + pm10no3)*adjust) +(co*.008);

* Calculate generic predictions from pmfine, pms4n3tc, pms4n3c and
pms4n3co;
* These are the same regression equations whose rc are in PM25A2RC.SD2;
* Calculate predicted PM2.5 here so that QA checks can be applied
before selecting best predictor;

pred_TCg=((pmsntcg*1.76409) + 1.66990);
pred_Cg=((pms4n3cg*1.75881) - 0.63789);
pred_COg=((pmsncog*1.56442) + 0.7441);
pred_di=((pmfine2*1.16153) - 0.01683);

*QA for PM predictors;
difTC=pm10nat- pred_TCg;
difC=pm10nat - pred_Cg;
difco=pm10nat - pred_COg;

* Use -10 since actual PM2.5 can be as much as PM10, +/- some msmt
error on both;
If(diftc lt -10 and pm10nat ne .) then do;
  pmsntcg=.;
  pms4n3tc=.;
  end;
If(difc lt -10 and pm10nat ne .) then do;
  pms4n3cg=.;
  pms4n3c=.;
  end;
If(difco lt -10 and pm10nat ne .) then do;
  pmsncog=.;
  pms4n3co=.;
  end;


************* Prioritize data for site specific predictions
****************;
predictr=pm25nat;
if(pm25nat eq .) then predictr=pmfine;
if(pm25nat eq . and pmfine eq .) then predictr=pm25spst;
if(pm25nat eq . and pmfine eq . and pm25spst eq .) then
predictr=pms4n3tc;
if(pm25nat eq . and pmfine eq . and pm25spst eq . and pms4n3tc eq .)
then predictr=pms4n3c;
if(pm25nat eq . and pmfine eq . and pm25spst eq . and pms4n3tc eq . and
pms4n3c eq .)
                                            then predictr=pms4n3co;

************* Mark source of best estimate ******************;
if(pm25nat ne .) then source='PM25NAT ';
if(pm25nat eq .) then source='PMFINEDI';
if(pm25nat eq . and pmfine eq .) then source='PM25SPST';
if(pm25nat eq . and pmfine eq . and pm25spst eq .) then
source='PMS4N3TC';
if(pm25nat eq . and pmfine eq . and pm25spst eq . and pms4n3tc eq .)
then source='PMS4N3C ';
if(pm25nat eq . and pmfine eq . and pm25spst eq . and pms4n3tc eq . and
pms4n3c eq .)
```

```
                                                                     then
source='PMS4N3CO';
if(pm25nat eq . and pmfine eq . and pm25spst eq . and pms4n3tc eq . and
pms4n3c eq .
    and pms4n3co eq . ) then source='LTSCAV24';
if(pm25nat eq . and pmfine eq . and pm25spst eq . and pms4n3tc eq . and
pms4n3c eq .
    and pms4n3co eq . and ltscav24 eq .) then source='          ';


************** Remove observations with no pm25nat and no data to use
for predictions *******;
if(source ne '          ');
proc sort;
  by site source;


***************************************************;

*************** Get regression coefficients to predict pm25
***************;
data tempcorc;     *rc from co by site;
set sasfiles.pmCOsirc;
source=species;


data temp1;
set sasfiles.pm25parc;      *site specific reg coef;
IF(DEPENDNT eq 'PM25NAT '); *keep only regression coefficients for
predicting pm25nat (FRM);
source=indepent;

data tempboth;
set tempcorc temp1;
slopesi=slope;
intersi=intercep;
nxsi=nx;  mxsi=mx;  sxxsi=sxx;  S2SI=S2;
*check for r2 and number of observations;
if(r2 lt .8  or nx lt 30) then do; * only use site-specific regressions
if good r2;
slopesi=.;
intersi=.;
  end;

proc sort;
  by site source;


*************** Merge site specific rc with daily data
*********************;
data temp3;
merge tempdat4 tempboth;
  by site source;

keep site year month day pmfine pmfine2 pm25nat pm10nat pm25spst
s2predsp  pms4n3
     pms4n3tc pms4n3c  pms4n3co LTSCAV24 adjust elev  pmsntcg  pms4n3cg
pmsncog
```

```
      nxSI mxSI sxxSI SLOPEsi INTERsi S2SI source PREDICTR pred_di
pred_TCg
      pred_Cg pred_COg;
*keep both unadjusted and elevation-adjusted data--will use based on
whether or not
  the program picks site specific or generic rc;
proc sort;
  by source;

 ************* Create permanent site-specific prediction file
*****************;

data sasfiles.dailyss;
set temp3;

*************************************************************************
********;

************* Acquire generic rc ********************;
data tempgen;
set sasfiles.pm25a2rc;
if(indepent eq 'PMFINEDI') then source='PMFINEDI';
IF(indepent eq 'PMS4N3C ') THEN SOURCE='PMS4N3C ';
if(indepent eq 'PMS4N3TC') Then source='PMS4N3TC';
IF(indepent eq 'PMS4N3CO') then source='PMS4N3CO';
slopeG=slope;
interG=intercep;
nxg=nx; mxg=mx;  sxxg=sxx; S2G=S2;
keep source slopeG interG nxG mxG SxxG S2G;

proc sort;
  by source;

***************************************************;

********** This section substitutes predictions from generic
regressions
         when site-specific regressions are unavailable
****************;

data tempdat5;
merge temp3 tempgen ;
  by source;

*correct predictr if there is no site specific rc;
if(slopeSI eq .) then do;
*prioritize data for predictions when there is no site specific rc;
predictr=pm25nat;
if(pm25nat eq .) then predictr=pmfine2;
if(pm25nat eq . and pmfine2 eq .) then predictr=pm25spst;
if(pm25nat eq . and pmfine2 eq . and pm25spst eq .) then
predictr=pmsntcg;
if(pm25nat eq . and pmfine2 eq . and pm25spst eq . and pmsntcg eq .)
then predictr=pms4n3cg;
if(pm25nat eq . and pmfine2 eq . and pm25spst eq . and pmsntcg eq . and
pms4n3cg eq .)
                                        then predictr=pmsncog;
```

```sas
if(pm25nat eq . and pmfine2 eq . and pm25spst eq . and pmsntcg eq . and
pms4n3cg eq .
    and pmsncog eq .) then predictr=ltscav24;


*mark source;
if(pm25nat ne .) then source='PM25NAT ';
if(pm25nat eq .) then source='PMFINEDG';
if(pm25nat eq . and pmfine2 eq .) then source='PM25SPST';
if(pm25nat eq . and pmfine2 eq . and pm25spst eq .) then
source='PMSNTCG ';
if(pm25nat eq . and pmfine2 eq . and pm25spst eq . and pmsntcG eq .)
then source='PMS4N3CG';
if(pm25nat eq . and pmfine2 eq . and pm25spst eq . and pmsntcG eq . and
pms4n3cG eq .)
                                                            then
source='PMSNCOG ';
if(pm25nat eq . and pmfine2 eq . and pm25spst eq . and pmsntcG eq . and
pms4n3cG eq .
    and pmsncoG eq .) then source='LTSCAV24';
if(pm25nat eq . and pmfine2 eq . and pm25spst eq . and pmsntcG eq . and
pms4n3cG eq .
    and pmsncoG eq . and ltscav24 eq .) then source='        ';
*remove data with no pm25nat and no data to use for predictions;
if(source ne '        ');
  END;
proc sort;
  by site source;


*********************************************************************
********;

****** Calculate predictions from pm25nat, pm25spst and ltscav24
**************;


data temp2;
set tempdat5;

pred_nat=pm25nat;
pred_sps=pm25spst;
bsp=(100*ltscav24) - 10;
dPM25bsp = 0.14*bsp;
s2prdbsp=((1+(1/979))*(77023/(979-2)) ) + ( ((164.72-bsp)**2) *
((0.00131)**2) );
pred_bsp=-0.017 + (1.16*dpm25bsp*adjust);  *adjusts for elevation;

************* Calculate best predictions *******************;
if(source eq 'PM25NAT ') then do;
   predictd=predictr;
   s2pred=0;
   end;
if(source eq 'PM25SPST' ) THEN DO;
   predictd=predictr;
   s2pred=s2predSP;
   end;
```

```
 *************;if(source eq 'LTSCAV24' ) then do;
  *there will only be ltscav24 for years before 1995 -- later years
were removed at
    the beginning of this program;
if(bsp ge 800) then do;
  predictd=.;
  s2pred=.;
end;
if(bsp lt 800) then do;
*n=979, SSE=77,023, xbar=164.72, Seslope=0.00131 ;
predictd =-0.017 + (1.16*dpm25bsp*adjust);  *adjusts for elevation;
s2pred=( (1+(1/1053))*(14842.754/1051) ) + ( ((16.661-
(dpm25bsp*adjust))**2)*(0.007439**2) );
end;
************; end;
************For site specific RC;
if(source ne 'PM25NAT ' and source ne 'PM25SPST' and source ne
'LTSCAV24' AND slopesi ne .)
              then do;
  predictd=(slopeSI*predictr) + interSI;
  s2pred=(s2SI*(1 + (1/nxSI) + ( ((predictr-mxSI)**2)/SxxSI)));
  end;

***********For generic RC;
  if(source ne 'PM25NAT ' and source ne 'PM25SPST' AND slopesi EQ . and
source ne 'LTSCAV24' )
       then do;
  predictd=(slopeG*predictr) + interG;
  s2pred=(s2G*(1 + (1/nxG) + ( ((predictr-mxG)**2)/SxxG)));

  end;

********** Make permanent SAS data set ***************;
data sasfiles.dailysga;
set temp2;

*****************************************************;

data tempday;
set temp2;
if(predictd ne .);  *keep only obs with predicted data;
keep site year month day source predictr predictd s2pred pred_nat
pred_di
     pred_bsp pred_sps pred_TCg pred_Cg pred_COg;

proc sort;
  by site year month source;



proc means noprint data=tempday;   *means for each month, by sources;
  var predictd ;
  by site year month source;
  output out=tempmnsr
  mean=predictd
  n=npred;
```

```
********************************************************;

****************** Set up file with monthly means by source ***********;
data temp5;
set tempmnsr;
proc sort;
  by site year month;
proc transpose out=temp6;       *temp6 has the number of times source
occurs in one month;
  var npred;
  by site year month;
  id source;

proc means noprint data=tempday;    *means for each month, all sources;
  var predictd s2pred pred_nat pred_di pred_sps pred_TCg pred_Cg
pred_COg pred_bsp;
  by site year month;
  output out=tempmn
  mean=predictd s2pred pred_nat pred_di pred_sps pred_TCg pred_Cg
pred_COg pred_bsp
  n=npred ns2pred nprd_nat npred_di nprd_sps nprd_TCg npred_Cg nprd_COg
nprd_bsp
  std=sdpred sds2pred sdpr_nat sdprd_di sdpr_sps sdpr_TCg sdprd_Cg
sdpr_COg sdpr_bsp;

*********************************************************;

************* Add in TWS monthly predictions *******************;
data tempsp2;
set sasfiles.spstdtws;
if(year lt 2003);
pred_TWS=pm25;
s2prdTWS=s2pred;
keep site year month  pred_TWS s2prdTWS;
proc sort;
  by site year month;

data temp7;
merge tempmn tempsp2;
  by site year month;

****************************************************************;
************* Create final output file ******************;

data sasfiles.eachMn4G;
merge temp7 temp6;
  by site year month;

*use tws only when there is no daily predictor available;
if(predictd eq .) then do;
  predictd=pred_tws;
  s2pred=s2prdtws;
  sdpred=0;
  npred=1;
  end;
************************;
```

```
uncert=sqrt( ((sdpred**2) + s2pred)/npred );

************* Flag Deviations **************************;
devFINE=abs(predictd-pred_di);
devSNTC=abs(predictd-pred_TCg);
devS4N3C=abs(predictd-pred_Cg);
devSNCO=abs(predictd-pred_COg);
devltsct=abs(predictd-pred_bsp);
devspst=abs(predictd-pred_sps);
devtws=abs(predictd-Pred_tws);
devpm25n=abs(predictd-pred_nat);


maxdev=max(devfine, devsntc, devs4n3c, devsnco, devspst, devpm25n);
sigma2x=2*uncert;

if(maxdev gt 10 and maxdev gt sigma2x ) then flagmxdv=1;
if(devltsct gt 10 and devltsct gt sigma2x) then flagneph=1;
if(devtws gt 10 and devtws gt sigma2x) then flagtws=1;



*add info about primary source for data for each month;
*select most frequent source for each month;
if(pm25nat ge pmfinedi and pm25nat ge pmfinedg and pm25nat ge pms4n3tc
and pm25nat ge pmsntcg
    and pm25nat ge pms4n3c and pm25nat ge pms4n3cg and PM25nat ge
ltscav24
    and pm25nat ge pms4n3co and pm25nat ge pmsncog and pm25nat ge
pm25spst )
            then source1='pm25nat ';

if(pmfinedi ge pm25nat and pmfinedi ge pmfinedg and  pmfinedi ge
pms4n3tc
    and pmfinedi ge pmsntcg and pmfinedi ge pms4n3c and pmfinedi ge
pms4n3cg
    and pmfinedi ge pms4n3co and pmfinedi ge pmsncog and pmfinedi ge
pm25spst
    and pmfinedi ge ltscav24)
            then source1='pmfinedi';

if(pmfinedg ge pm25nat and pmfinedg ge pmfinedi and  pmfinedg ge
pms4n3tc
    and pmfinedg ge pmsntcg and pmfinedg ge pms4n3c and pmfinedg ge
pms4n3cg
    and pmfinedg ge pms4n3co and pmfinedg ge pmsncog and pmfinedg ge
pm25spst
    and pmfinedg ge ltscav24)
            then source1='pmfinedg';

if(pms4n3tc ge pmfinedi and pms4n3tc ge pmfinedg and  pms4n3tc ge
pm25nat and pms4n3tc ge pms4n3cg
 and pms4n3tc ge pms4n3c and pms4n3tc ge pmsncog and pms4n3tc ge
pmsntcg
    and pms4n3tc ge pms4n3co and pms4n3tc ge pm25spst and pms4n3tc ge
ltscav24)
            then source1='pms4n3tc';
```

```
if(pmsntcg ge pmfinedi and pmsntcg ge pmfinedg and  pmsntcg ge pm25nat
and pmsntcg ge pms4n3cg and
 pmsntcg ge pms4n3c and pmsntcg ge pmsncog  and pmsntcg ge pms4n3tc
   and pmsntcg ge pms4n3co and pmsntcg ge pm25spst and pms4n3tc ge
ltscav24 )
          then source1='pmsntcg';

if(pms4n3c ge pmfinedi and pms4n3c ge pmfinedg and pms4n3c ge pm25nat
and pms4n3c ge pms4n3tc
   and pms4n3c ge pmsntcg and pms4n3c ge pms4n3co and pms4n3c ge
pmsncog and pms4n3c ge pm25spst
   and pms4n3c ge pms4n3cg and pms4n3c ge ltscav24)
          then source1='pms4n3c ';

if(pms4n3cg ge pmfinedi and pms4n3cg ge pmfinedg and pms4n3cg ge
pm25nat and pms4n3cg ge pms4n3tc
   and pms4n3cg ge pmsntcg and pms4n3cg ge pms4n3co and pms4n3cg ge
pmsncog and
 pms4n3cg ge pm25spst and pms4n3cg ge pms4n3c and pms4n3cg ge ltscav24)
          then source1='pms4n3cg';

if(pms4n3co ge pmfinedi and pms4n3co ge pmfinedg and  pms4n3co ge
pm25nat
   and pms4n3co ge pms4n3tc and pms4n3co ge pmsntcg and pms4n3co ge
pmsncog
   and pms4n3co ge pms4n3c and pms4n3co ge pms4n3cg and pms4n3co ge
pm25spst
   and pms4n3co ge ltscav24)
          then source1='pms4n3co';

if(pmsncog ge pmfinedi and pmsncog ge pmfinedg and  pmsncog ge pm25nat
   and pmsncog ge pms4n3tc and pmsncog ge pmsntcg and pmsncog ge
pms4n3co
   and pmsncog ge pms4n3c and pmsncog ge pms4n3cg and pmsncog ge
pm25spst
   and pmsncog ge ltscav24)
          then source1='pmsncog ';

if(pm25spst ge pmfinedi and pm25spst ge pmfinedg and  pm25spst ge
pm25nat
   and pm25spst ge pms4n3c and pm25spst ge pms4n3cg and pm25spst ge
pms4n3tc
   and pm25spst ge pmsntcg and pm25spst ge pms4n3co and pm25spst ge
pmsncog
   and pm25spst ge ltscav24)
          then source1='pm25SPST';

if(ltscav24 ge pmfinedi and ltscav24 ge pmfinedg and  ltscav24 ge
pm25nat
   and ltscav24 ge pms4n3c and ltscav24 ge pms4n3cg and ltscav24 ge
pms4n3tc
   and ltscav24 ge pmsntcg and ltscav24 ge pms4n3co and ltscav24 ge
pmsncog
   and ltscav24 ge pm25spst )
          then source1='ltscav24';
```

```
*include tws as predictor only when all daily predictors are missing;
if(ltscav24 eq . and pmfinedi eq . and pmfinedg eq . and pm25nat eq .
   and pm24n3c eq . and pms4n3cg eq . and pm24n3tc eq .
   and pmsntcg eq . and pms4n3co eq . and pmsncog eq .
   and pm25spst eq . and pred_tws ne .) then source1='spstdtws';


run;

********************* END OF PROGRAM *********************;
```

APPENDIX I. MONITORING SITES WITH INCOMPLETE SITE INFORMATION

We obtained information on site locations and elevations from CARB. Missing data (including elevation) were recorded as zeros in the CARB information file. Few of the sites with incomplete location or elevation information reported PM data for more than one year. We also noted two sites with apparently incorrect elevations (neither reported PM measurements).

Table I1. List of monitoring sites with incomplete information on location or elevation.

Code	Site Name	Latitude	Longitude	Elevation (m)	PM Start Year	PM End Year	No. PM Months
2011	Eureka-Fort Avenue	40.802	-124.163	0			
2037	Eureka-Myrtle Avenue	40.802	-124.163	0			
2046	Merced-Merced College	37.302	-120.482	0			
2048	Middletown	38.753	-122.614	0			
2049	Davis-Brown Drive	38.545	-121.739	0			
2061	Sequoia-Lookout Point	36.429	-118.768	0			
2106	Bishop-Main	37.363	-118.397	4120			
2111	Quincy-County Courthouse	39.937	-120.946	0			
2139	Quincy-S Redburg Avenue	39.937	-120.946	0	1986	1987	12
2154	South Lake Tahoe-Main Post Office	38.946	-119.970	0			
2158	Burbank-Monterey Avenue	34.180	-118.330	0			
2173	El Centro-Broadway	32.793	-115.438	0			
2177	Maricopa-Ozena Station	35.059	-119.400	0			
2187	Rialto-Airport	34.106	-117.369	0			
2189	Lancaster-N Cedar Avenue	34.698	-118.136	0			
2202	Lone Pine-Visitor Center	36.606	-118.062	0			
2203	Piru-Temescal Station	34.415	-118.793	0			
2226	Bakersfield-Rio Bravo	35.373	-119.018	0			
2231	Little Lake	35.937	-117.906	0			
2232	Sacramento-Cal Expo/Am Youth Hostel	38.582	-121.493	0			
2237	Pacifica-San Pedro	37.614	-122.486	0			
2259	Sacramento-County Ag Office	38.582	-121.493	0			
2269	El Centro-State Street	32.792	-115.435	0			
2281	South Lake Tahoe-Blackbart	38.946	-119.970	0			
2302	Gridley-Graylodge	39.364	-121.693	0			
2304	Sacramento-Florin	38.582	-121.493	0			
2326	Fontana-Redwood	34.092	-117.434	0			
2330	Sutter Creek-Main Street	38.393	-120.801	0			
2331	Mammoth Lakes-Sherwin Creek Road	37.649	-118.971	0			
2334	Carson-Victoria Street	33.831	-118.281	0			
2361	Santa Barbara-Los Prietos	34.423	-119.703	0			

Code	Site Name	Latitude	Longitude	Elevation (m)	PM Start Year	PM End Year	No. PM Months
2362	Bishop-S Main Street	37.361	-118.393	0			
2366	Coso Junction-10 miles E	36.034	-117.799	0			
2378	South Lake Tahoe-CalTrans Yard	38.946	-119.970	0			
2387	Weaverville-Hospital	40.677	-122.939	0	1987	1987	9
2396	Mountain Home-SF Headquarters	34.101	-116.998	0			
2398	Santa Maria-Lake Marie East	34.953	-120.435	0			
2401	Westmoreland-Route 86	33.119	-115.184	0			
2407	El Cajon	32.795	-116.962	0			
2424	East Biggs	39.415	-121.653	0	1982	1983	4
2436	Darwin-Quintana Office	36.268	-117.591	0			
2437	Crows Landing-Davis	37.371	-121.132	0			
2438	Mammoth Lakes-4 miles SE	37.649	-118.971	0			
2445	Quincy-Fairgrounds	39.942	-120.917	0			
2459	Sacramento-1131 S Street	38.582	-121.493	0			
2468	San Juan Capistrano	33.502	-117.662	0			
2479	Whispering Pines	38.814	-122.711	0			
2482	La Conchita-7128 Santa Paula	34.371	-119.306	0			
2495	Long Beach-San Antonio Drive	33.767	-118.188	0			
2498	Modesto-Jennings Road	37.639	-120.996	0			
2505	Elk Grove-Via Media	38.409	-121.371	0			
2517	Yosemite-Camp Six	37.547	-119.842	0			
2520	Geyserville-Redwood Freeway	38.708	-122.901	0			
2522	Weaverville-CalTrans	40.731	-122.941	0			
2533	Willow Creek-CSD Highway 96	40.940	-123.630	0			
2540	Big Bear Lake	34.244	-116.911	0			
2550	Los Olivos-Figueroa Station	34.668	-120.114	0			
2554	Valley Home-School	37.829	-120.911	0			
2561	Brawley-Hovely	33.019	-115.461	0			
2568	Kelseyville-Kelsey Creek Drive	38.978	-122.838	0			
2569	San Diego-Front	32.716	-116.836	0			
2595	Anderson-Kimberly Road	40.448	-122.297	0			
2599	Vandenberg Air Force Base-Pt Arguello #1	34.684	-120.603	0			
2610	Quincy-CHP Building	39.934	-120.942	0			
2624	Victoria-1000 Victoria #2	34.052	-118.243	0			
2652	Fontana-Cypress	34.092	-117.434	0			
2654	Calexico-Fire Station	32.660	-115.490	0			
2669	San Bernardino	34.121	-117.302	0			
2670	Santa Ana-Police Station	33.746	-117.867	0			
2688	Sacramento-Metro Airport Tower	38.582	-121.493	0	1983	1985	27
2695	Oroville-County Center	39.514	-121.555	0			
2721	Little Lake-Highway 395	35.937	-117.906	0			
2733	Saratoga-Highway 85 and SP RR	37.264	-122.022	0			
2736	Ventura-Casitas Station	34.371	-119.306	0			
2738	Hayfork-Ranger Station	40.554	-123.182	0			
2742	Maricopa-Ventura Station	35.059	-119.400	0			

Code	Site Name	Latitude	Longitude	Elevation (m)	PM Start Year	PM End Year	No. PM Months
2746	Sutter-County Yard	39.160	-121.748	0			
2762	Riverside-RCC	33.953	-117.395	0			
2781	Santa Ana-Weir Canyon Road	33.746	-117.867	0			
2783	Arbuckle-Lucas Street	39.017	-122.057	0			
2786	Fillmore-Oak Flat Station	34.399	-118.917	0			
2791	Madison-Main Street	38.679	-121.967	0			
2794	Peppermint-Heliport	36.208	-119.346	0			
2818	Broderick-3rd Street	38.591	-121.516	0			
2841	Cobb-Binkley Ranch	38.822	-122.722	0			
2850	Berkeley	37.870	-122.270	0			
2853	Independence-10 miles N-Blackrock	36.803	-118.199	0			
2885	Calipatria-6.5 miles NW	33.178	-115.390	0			
2887	Biggs-9th and C Street	39.403	-121.719	0			
2889	Dunnigan-Rest Area I5 East	38.885	-121.969	0	1982	1984	28
2890	Laguna Beach-Arroyo	33.542	-117.782	0			
2900	Mammoth Lakes-Water District	37.649	-118.971	0			
2928	Burney-High School	40.884	-120.351	0			
2934	Commerce-Indiana & Shelia	34.007	-117.809	0			
2951	Ramona-Airport Road	33.042	-116.867	0			
2975	Industry-7th	34.042	-117.961	0			
2976	Carson-Sherman	33.831	-118.281	0			
2982	Scotts Valley-Vine Hill	37.060	-122.000	0			
2989	Commerce-Ayers	34.005	-117.824	0			
3012	Industry-Salt Lake	34.041	-117.961	0			
3031	La Jolla-Mount Soledad	33.858	-117.876	0			
3131	Crescent City-9th and H Street	41.764	-124.200	0			
3139	Tijuana-Center of Health #1	32.527	-117.039	0	1996	1997	23
3142	Winterhaven-2nd Avenue	32.739	-114.636	0	1996	1996	2
3148	Davis-Russel Blvd	38.546	-121.745	0			
3184	Mexicali-Profepa	32.615	-115.436	0	1996	1996	4
3185	Mexicali-ITM	32.621	-115.398	0	1996	1999	37
3189	Blue Lake-Greenwood Avenue	40.885	-123.991	0			
3190	Mexicali-Museo	32.666	-115.454	0	1996	1996	3
3191	Mexicali-Odontolgia	32.632	-115.453	0	1996	1996	4
3192	Mexicali-Buen Pastor	32.680	-115.430	0	1996	1996	4
3193	Mexicali-Conalep	32.570	-115.349	0	1997	1997	11
3204	Mexicali-Progreso	32.582	-115.584	0	1997	1997	10
3206	Sequoia National Park-Lookout Point	36.429	-118.763	3885			
3224	Paradise	39.770	-121.600	0			
3225	Antioch	0.000	0.000	0			
3226	Martinez	0.000	0.000	0			
3227	Walnut Creek	0.000	0.000	0			
3483	Stateline-Harveys Hotel	38.956	-119.945	0			

APPENDIX J. PROJECT REVIEW AND QUALITY ASSURANCE

Project Review

Developing a reliable historical record of fine PM mass concentrations necessitates combining data from different monitoring programs, accounting for differences in measurement methods and accuracy. Measurements of fine PM mass concentrations were used whenever available; when fine mass measurements were not available, they were reconstructed from related measurements, such as light scattering. We compared different types of measurements to identify inconsistencies (Level 3 validation) and used these comparisons to identify and exclude suspect reconstructions.

For each monthly estimated concentration, we generated an accompanying uncertainty estimate (in $\mu g\ m^{-3}$), which reflected both sampling and estimation uncertainties. The error analysis (assessment of possible biases, or systematic errors) methodology was reviewed by ARB staff with expertise in statistical methods for characterizing estimation errors.

We carried out most of the manipulation of databases using SAS. Appendix H provides complete documentation of the computer programs. The SAS code was reviewed and approved by a statistician from Research Division with expertise in SAS computer programming. The statistician determined that the thoroughly documented SAS code would correctly execute the functions and routines described by the investigators.

During the course of this research contract, we held several conference calls with ARB staff and health scientists to discuss and evaluate work in progress, and to plan work before the next progress report or conference call. The first version of the report and database were reviewed by ARB staff, staff from the Office of Environmental Health Hazard Assessment and California Department of Health Service, and Professor Michael Jerrett of the University of Southern California. The final version incorporates all reviewer comments.

Quality Assurance (QA) and Quality Control (QC)

Quality assurance and quality control encompass those activities that complement the measurement process by providing estimates of data accuracy, precision, validity, and representativeness, and ensuring that these attributes lie within acceptable limits. A rigorous quality assurance/quality control (QA/QC) program is a critical element for the success of any research project. The quality of the PM data used in this research study has been described in various ARB or district final reports or in numerous peer-reviewed publications.

All measurements used in this project were obtained from ARB archives, which consist of data that have been reviewed by ARB staff. All such data are considered valid. However, measurements of PM mass concentrations may differ among types of samplers, none of which is considered to be an absolute standard. We used the EPA Federal Reference Method (FRM) sampler as a standard, adjusting all other measurements of fine PM mass concentrations to the equivalent FRM values.

The historical record of fine PM mass concentrations combined data from different monitoring programs, accounting for differences in measurement methods and accuracy. Data quality (accuracy, uncertainty, completeness, etc.) and the degree of correlation with FRM fine PM mass concentrations varied among alternative and reconstructed mass measures. All measured and reconstructed fine PM mass concentrations exhibited a correlation with FRM fine mass concentrations of r^2 equal to, or exceeding, 0.8.

The principal sources of the uncertainties of station averages are: (1) measurement accuracy and precision, (2) uncertainty associated with use of reconstructed mass measurements, (3) uncertainty associated with incomplete sampling (e.g., four measurements per month), and (4) uncertainty associated with estimation from other locations using intersite correlations (however, the database that we prepared did not include estimates derived by spatial interpolation). For each station and each month, the best estimate of the monthly fine mass average concentration was accompanied by an

uncertainty estimate, explanatory information describing which data were used in the estimation, and monthly averages constructed from a variety of measurement methods for comparison with the best estimates.

In developing monthly averages of measured and reconstructed fine mass constructions, we established a selection priority as follows:

1. FRM fine mass
2. dichotomous sampler fine mass
3. CADMP fine mass and fine mass from other special studies
4. reconstruction from PM10 sulfate + nitrate + total carbon
5. reconstruction from PM10 sulfate + nitrate + total carbon calculated from CoH
6. reconstruction from PM10 sulfate + nitrate + total carbon calculated from CO
7. reconstruction from nephelometer data prior to 1995

Other reconstructions of fine PM mass concentrations did not exhibit a correlation with FRM fine mass concentrations of r^2 equal to, or exceeding, 0.8, and were therefore not used. For each day of a month, a daily-average PM level was obtained following the preceding priorities. Then, a monthly average was determined from all days in a month having data.

In this study, methods for reconstructing fine PM mass concentrations were based on established principals from earlier work. Previous studies have shown that the principal constituents of $PM_{2.5}$ mass in California are organic and black (elemental) carbon, sulfate, and nitrate. These PM components, in turn, are typically found primarily in the fine fraction. As a result, it was possible to reconstruct fine mass concentrations and their uncertainties at places and during times without measurements of $PM_{2.5}$ mass using measurements of sulfate, nitrate, and carbon from PM_{10} samples. This method of reconstructing fine mass concentrations may be compared with procedures used by the Interagency Monitoring of Protected Visual Environments (IMPROVE) program, whose

results, in general, indicate that the sum of the fine composites provides a reasonable estimate of the fine mass concentration (IMPROVE, 2002).

Reconstruction of fine mass concentrations from light scattering data has also been examined in previous studies. For example, Husar and Falke (1996) conducted a comparative study of aerosol light scattering and fine particle mass data for fourteen different sites in the western U.S. (including six sites in California). The data for the fourteen sites indicated a good correlation, with half of the sites exhibiting R^2 above 0.8. Groblicki et al. (1981) presented the light scattering coefficient observed in studies in Denver, Colorado as a function of the observed mass in the fine and coarse particle ranges, respectively. It has been seen that a good linear relationship exists between scattering coefficient and the fine mass, but not between scattering coefficient and coarse particle mass. Our own previous research work indicated that estimation of fine PM mass concentrations from the ARB nephelometer network requires careful investigation of calibration changes over time. In addition, in the present study, several comparisons were made with other data (e.g., relative humidity) to exclude nephelometer measurements potentially influenced by fog or cloud droplets.

In reconstructing fine PM mass concentrations, we employed several comparisons of different types of measurements to help identify and exclude outliers. For example, predictions of fine mass concentrations that exceeded measured values of total suspended particulate (TSP) were deemed invalid. Predictions of fine mass concentrations that exceeded measured values of PM_{10} mass by more than 10 µg m^{-3} were also deemed invalid (providing an allowance for uncertainties in both reconstructed fine mass and measured PM10 mass concentrations). Similarly, reconstructions of carbon mass concentrations from related measurements, including coefficient of haze (CoH) and carbon monoxide (CO) were compared with PM_{10} mass concentrations less measured levels of sulfate, nitrate, and ammonium to check for consistency.

REFERENCES

Groblicki PJ, Wolff GT, and Countess RJ. 1981. Visibility reducing species in the Denver brown cloud. Atmos. Environ. 15: 2473-2484.

Husar RB and Falke R. 1996. The relationship between aerosol light scattering and fine mass. Report No. CX 824179-01; prepared for Office of Air Quality Planning and Standards, U.S. EPA, Research Triangle Park NC.

IMPROVE. 2002. Interagency Monitoring of Protected Visual Environments—Data Resources. National Park Service, Ft. Collins CO. http://vista.cira.colostate.edu/IMPROVE.

Kokoska S. and Nevison C. 1989. Statistical Tables and Formaulae. Springer-Verlag, New York. 88 pp.

McDade C. 1997. IMS95 Data Analysis. CARB contract number 97-7PM.

McMurry P., Shepherd M., and Vickery J. 2004. Particulate Matter Science for Policy Makers: A NARSTO Assessment. Cambridge University Press, New York. 510 pp.

Motallebi N., Taylor Jr. C., Turkiewicz K., and Croes B. 2003. Particulate matter in California: Part 1 – Intercomparison of several $PM_{2.5}$, $PM_{10-2.5}$, and PM_{10} monitoring networks. J. Air Waste Manage. Assoc. 53: 1509-1516.

Motallebi N., Taylor Jr. C., and Croes B. 2003. Particulate matter in California: Part 2 – Spatial and compositional patterns of $PM_{2.5}$, $PM_{10-2.5}$, and PM_{10}. J. Air Waste Manage. Assoc. 53: 1509-1516.